

Annual Technical Report for ACCESS for ELLs Online English Language Proficiency Test

Series 602, 2023-2024 Administration
Annual Technical Report No. 20A

Part 2: Technical Results

Prepared by Center for Applied Linguistics
Language Assessment Division
Psychometrics and Quantitative Research Team
June 2025

Contents

1. Annual Test Results	7
1.1 Participation	9
1.1.1 Grade-Level Cluster	9
1.1.2 Grade.....	12
1.2 Scale Score Results	17
1.2.1 Mean Scale Score Across Domain and Composite Score by Cluster.....	17
1.2.2 Mean Scale Score Across Domain and Composite Score by Grade	23
1.2.3 Correlations.....	39
1.3 Proficiency Level Results	41
1.3.1 Domains	41
1.3.2 Composites	47
2. Analysis of Domains	54
2.1 Complete Item or Task Analysis and Summary	56
2.1.1 Listening	60
2.1.2 Reading	70
2.1.3 Writing	85
2.1.4 Speaking	95
2.2 DIF Analysis and Summary	99
2.2.1 Listening	102
2.2.2 Reading	105
2.2.3 Writing	107
2.2.4 Speaking	109
2.3 Raw Score Distribution.....	113
2.3.1 Listening	113
2.3.2 Reading	113
2.3.3 Writing	114
2.3.4 Speaking	124
2.4 Scale Score Distribution	139

2.4.1	Listening	140
2.4.2	Reading	145
2.4.3	Writing	150
2.4.4	Speaking	165
2.5	Proficiency Level Distributions.....	185
2.5.1	Listening	186
2.5.2	Reading	191
2.5.3	Writing	196
2.5.4	Speaking	211
2.6	Raw Score to Scale Score to Proficiency Level Conversion for Speaking and Writing	231
2.6.1	Listening	231
2.6.2	Reading	231
2.6.3	Writing	232
2.6.4	Speaking	242
2.7	Equating Summary	251
2.7.1	Listening	258
2.7.2	Reading	268
2.7.3	Writing	279
2.7.4	Speaking	289
2.8	Test Characteristic Curve.....	294
2.8.1	Listening	295
2.8.2	Reading	295
2.8.3	Writing	295
2.8.4	Speaking	300
2.9	Test Information Function.....	308
2.9.1	Listening	311
2.9.2	Reading	313
2.9.3	Writing	316

2.9.4	Speaking	323
3.	Analysis of Composite Scores	334
3.1	Scale Score Distribution for Composites.....	334
3.1.1	Oral	335
3.1.2	Literacy	340
3.1.3	Comprehension	345
3.1.4	Overall	350
3.2	Proficiency Level Distribution for Composites	355
3.2.1	Oral	356
3.2.2	Literacy	361
3.2.3	Comprehension	366
3.2.4	Overall	371
4.	Annual Updates of Validity Evidence.....	376
4.1	Standards.....	377
4.1.1	Test Content.....	377
4.1.2	Response Processes	377
4.1.3	Internal Structure	377
4.1.4	Relation to Other Variables.....	377
4.2	Annual Validity Studies	377
4.2.1	Validating a New Writing Scoring Scale Using Multi-Faceted Rasch Analysis	377
4.2.2	Development of a New WIDA Writing Scoring Rubric for Grades 1-12.....	378
4.2.3	Examining English Learner Testing, Proficiency, and Growth: Before, During, and "After" the COVID-19 Pandemic	380
5.	Reliability.....	382
5.1	Reliabilities of the Domain Scores.....	386
5.1.1	Listening	390
5.1.2	Reading	391
5.1.3	Writing	393
5.1.4	Speaking	396

5.2 Interrater Agreement Rates	399
5.2.1 Listening	400
5.2.2 Reading	400
5.2.3 Writing	401
5.2.4 Speaking	403
5.3 Conditional Standard Errors of Measurement of the Domain Scale Scores	407
5.3.1 Listening	409
5.3.2 Reading	412
5.3.3 Writing	415
5.3.4 Speaking	418
5.4 Accuracy and Consistency of Domains	421
5.4.1 Listening	427
5.4.2 Reading	428
5.4.3 Writing	430
5.4.4 Speaking	431
5.5 Reliabilities of Students' Composite Scale Scores	432
5.5.1 Oral	435
5.5.2 Literacy	438
5.5.3 Comprehension	441
5.5.4 Overall	445
5.6 Conditional Standard Errors of Measurement of the Composite Scale Scores	451
5.6.1 Oral	453
5.6.2 Literacy	458
5.6.3 Comprehension	463
5.6.4 Overall	468
5.7 Accuracy and Consistency of Composites	473
5.7.1 Oral	476
5.7.2 Literacy	477
5.7.3 Comprehension	479

5.7.4 Overall	480
6. Quality Control.....	482
6.1 Content Development Quality Control	482
6.2 Test Administration Quality Control	484
6.3 Rater Quality Control	485
6.4 Score Reporting Quality Control.....	486
6.5 Data Forensic Quality Control	487

1. Annual Test Results

This section of the report provides an overview of students' participation, the distribution of students' scale scores, and the distribution of students' proficiency levels to see student performance of the ACCESS 602 administration. Results are presented, where appropriate, by grade-level cluster, grade, and tier (for Writing and Speaking), and also by state, by gender, and by race and ethnicity.

The analyses in this section follow the [U.S. Census Bureau's approach to reporting race and ethnicity](#), in which ethnicity is a binary category (Hispanic or non-Hispanic), with five categories for race (American Indian/Alaskan Native, Asian, Black/African American, Pacific Islander/Hawaiian, and White) that are not mutually exclusive. Thus, for example, Student A may be labeled as Hispanic for ethnicity and Asian for race, while Student B may be labeled as non-Hispanic for ethnicity and both American Indian/Alaskan Native and Black/African American for race. Students who are labeled Hispanic are included in the Hispanic (of any race) category, regardless of how many racial categories they are included in. Students who are identified in one racial category (e.g., Asian) who have not been identified as Hispanic are identified in only one racial category; if they are identified in more than one racial category and have not been identified as Hispanic, they are labeled non-Hispanic multiracial.

A subset of students was included in the descriptions of student participation and performance but were excluded from subsequent analyses, namely those students who were flagged as potentially having experienced test interruptions (that is, testing experiences that are outside of regular testing experiences). Using telemetry data, WIDA selected three variables that might potentially indicate interruption. The interruption indicators WIDA used are (1) longer than expected testing time, (2) number of appearances (e.g., more than one) of test items, and (3) number of log-ins. Records were flagged if they fell outside of established criteria for any of these three indicators. WIDA included students whose records were flagged as interrupted in the tables that describe participation in the assessment but excluded them from all subsequent analyses. Tables 1.1 through 1.4 summarize the numbers of students excluded from these analyses. On average, 4% to 12% of students were excluded in each cluster and domain.

Table 1.1**Students Excluded from Analysis Due to Test Interruptions in Listening Domain by Cluster**

Cluster	No. of Excluded Students	Total Students	Percent
1	22200	235874	9.41%
2-3	48012	485075	9.9%
4-5	39558	413008	9.58%
6-8	64057	499905	12.81%
9-12	52215	545897	9.56%
Total	226042	2179759	10.37%

Table 1.2**Students Excluded from Analysis Due to Test Interruptions in Reading Domain by Cluster**

Cluster	No. of Excluded Students	Total Students	Percent
1	12654	235874	5.36%
2-3	32738	485075	6.75%
4-5	38663	413008	9.36%
6-8	53051	499905	10.61%
9-12	56302	545897	10.31%
Total	193408	2179759	8.87%

Table 1.3**Students Excluded from Analysis Due to Test Interruptions in Speaking Domain by Cluster**

Cluster	No. of Excluded Students	Total Students	Percent
1	21396	235874	9.07%
2-3	39376	485075	8.12%
4-5	35148	413008	8.51%
6-8	51512	499905	10.3%
9-12	44321	545897	8.12%
Total	191753	2179759	8.8%

Table 1.4**Students Excluded from Analysis Due to Test Interruptions in Writing Domain by Cluster**

Cluster	No. of Excluded Students	Total Students	Percent
1	-	235,874	0%
2-3	-	485,075	0%
4-5	31,111	413,008	7.53%
6-8	30,455	499,905	6.09%
9-12	33,544	545,897	6.14%
Total	95,110	2,179,759	4.36%

1.1 Participation

Participation in ACCESS Online is shown in three ways: by grade-level cluster, by grade, and, for Writing and Speaking only, by tier.

1.1.1 Grade-Level Cluster

Table 1.1.1.1 shows participation across the 41 WIDA states and U.S. territories that participated in the ACCESS Online operational testing program in 2023–2024 by grade-level cluster. The 41 rows show the number of students in that grade-level cluster who took the test by state, and the final row shows the total number of participants across all 41 states and U.S. territories. The state with the largest number of students was Illinois. The state/territory with the smallest number of participants was Palau. The biggest cluster was grades 9–12. The abbreviations are as follows: DC, District of Columbia; DD, Department of Defense Education Activity; MP, Northern Mariana Islands; BI, Bureau of Indian Education; PW, Palau, and VI, Virgin Islands.

Table 1.1.1.1**Participation by Cluster by State, S602 Online**

State	Cluster 1	Cluster 2–3	Cluster 4–5	Cluster 6–8	Cluster 9–12	Total
AK	759	1796	1958	2643	3028	10184
AL	4745	9047	7715	10305	10145	41957
BI	290	596	697	869	495	2947
CO	9934	21028	16434	18909	20386	86691
DC	1049	2141	1779	1928	2072	8969
DD	737	1514	1205	1075	774	5305
DE	1645	3362	2938	3504	4034	15483
GA	16841	33306	27257	31608	31056	140068
HI	1505	3257	3280	4189	3952	16183
ID	1863	3927	3567	4043	4292	17692

State	Cluster 1	Cluster 2-3	Cluster 4-5	Cluster 6-8	Cluster 9-12	Total
IL	24421	54149	46549	59273	60599	244991
IN	9067	18567	17101	20915	22063	87713
KY	5137	9775	7888	8256	9993	41049
MA	12956	25123	18355	20326	26205	102965
MD	11441	24050	19056	21825	24221	100593
ME	521	1248	1096	1357	1737	5959
MI	9473	17957	16025	20175	25656	89286
MN	8085	17238	14228	15893	16058	71502
MO	4336	8563	7077	7604	7720	35300
MP	73	233	242	435	371	1354
MT	248	603	719	824	669	3063
NC	16645	28849	26979	35587	36872	144932
ND	451	986	873	993	978	4281
NH	553	1149	922	1047	1200	4871
NJ	14291	27847	22154	25729	30594	120615
NM	4259	9512	8057	12999	15549	50376
NV	5833	12577	11070	13897	16059	59436
OK	5412	13565	12020	15605	16854	63456
PA	9503	18516	15416	20309	24175	87919
PW	-	-	202	323	-	525
RI	1696	3263	2996	4025	5386	17366
SC	5304	10690	9207	12551	14637	52389
SD	852	1595	1234	1382	1502	6565
TN	8685	16506	12789	14436	15768	68184
UT	4856	10810	9993	13789	15037	54485
VA	13243	29446	24304	25919	29515	122427
VI	90	173	222	297	337	1119
VT	161	324	301	374	451	1611
WA	13654	30542	28415	30949	31000	134560
WI	5000	10695	10219	13217	13796	52927
WY	260	550	469	521	661	2461
Total	235874	485075	413008	499905	545897	2179759

Table 1.1.1.2 shows participation by grade-level cluster by gender across all 41 states and U.S. territories combined, while Table 1.1.1.3 shows participation by grade-level cluster by ethnicity across all 41 states and U.S. territories. The gender ratio was generally 39% female, 45.5% male and 15.5% missing gender information in clusters. About 64%–68% of participants were Hispanic across all clusters.

Table 1.1.1.2

Participation by Cluster by Gender, S602 Online

Cluster	Statistic	Gender F	Gender M	Gender Missing	Total
1	Count	95582	104011	36281	235874
1	% within cluster	40.52%	44.10%	15.38%	100.00%
2-3	Count	197508	215790	71777	485075
2-3	% within cluster	40.72%	44.49%	14.80%	100.00%
4-5	Count	160528	185980	66500	413008
4-5	% within cluster	38.87%	45.03%	16.10%	100.00%
6-8	Count	189227	230275	80403	499905
6-8	% within cluster	37.85%	46.06%	16.08%	100.00%
9-12	Count	205785	254559	85553	545897
9-12	% within cluster	37.70%	46.63%	15.67%	100.00%
Total	Count	848630	990615	340514	2179759
Total	% within cluster	38.93%	45.45%	15.62%	100.00%

Table 1.1.1.3

Participation by Cluster by Ethnicity, S602 Online

Cluster	Statistic	Hispanic	Non-Hispanic	Unknown	Total
1	Count	152560	68152	15162	235874
1	% within cluster	64.68%	28.89%	6.43%	100.00%
2-3	Count	317294	137900	29881	485075
2-3	% within cluster	65.41%	28.43%	6.16%	100.00%
4-5	Count	271446	110729	30833	413008
4-5	% within cluster	65.72%	26.81%	7.47%	100.00%
6-8	Count	337982	119111	42812	499905
6-8	% within cluster	67.61%	23.83%	8.56%	100.00%
9-12	Count	370109	127103	48685	545897
9-12	% within cluster	67.80%	23.28%	8.92%	100.00%
Total	Count	1449391	562995	167373	2179759
Total	% within cluster	66.49%	25.83%	7.68%	100.00%

Table 1.1.1.4 shows participation by grade-level cluster and tier for all Writing and Speaking forms. In both Writing and Speaking domains, cluster 1 had a higher percentage of Tier A than Tier B/C, while in other clusters, percentages of Tier A became smaller. Pre-A counts in Speaking were relatively small.

Table 1.1.1.4

Participation by Cluster by Tier by Domain, S602 Online

Cluster	Tier	Writing	Speaking
1	PA	N/A	15259
1	A	209689	118617
1	BC	26157	101996
1	Total	235846	235872
2-3	PA	N/A	30725
2-3	A	161262	153319
2-3	BC	323756	301028
2-3	Total	485018	485072
4-5	PA	N/A	12962
4-5	A	116225	82038
4-5	BC	296778	318005
4-5	Total	413003	413005
6-8	PA	N/A	31745
6-8	A	218025	116776
6-8	BC	281866	351369
6-8	Total	499891	499890
9-12	PA	N/A	38973
9-12	A	204580	233760
9-12	BC	341293	273143
9-12	Total	545873	545876

1.1.2 Grade

This section provides tables parallel to those in the previous section but broken out by grade rather than by grade-level cluster. Table 1.1.2.1 shows student counts by grade and state, while Table 1.1.2.2 shows student counts by grade and gender, and Table 1.1.2.3 by grade and ethnicity. The largest grade was grade 2, and the smallest was grade 12. Table 1.1.2.4 presents the percentages between Tier A and B/C and indicates that most grades showed higher counts in tier B/C forms except in Speaking and Writing grade 1.

Table 1.1.2.1**Participation by Grade by State, S602 Online [Grade = G]**

State	G 1	G 2	G 3	G 4	G 5	G 6	G 7	G 8	G 9	G 10	G 11	G 12	Total
AK	759	886	910	1043	915	816	899	928	953	808	750	517	10184
AL	4745	4729	4318	4037	3678	3362	3440	3503	3809	2967	2149	1220	41957
BI	290	304	292	359	338	344	257	268	126	143	111	115	2947
CO	9934	10902	10126	8864	7570	6254	6300	6355	6510	5583	4646	3647	86691
DC	1049	1091	1050	1027	752	604	643	681	901	576	396	199	8969
DD	737	772	742	703	502	415	346	314	280	205	183	106	5305
DE	1645	1740	1622	1603	1335	1125	1165	1214	1510	1178	813	533	15483
GA	16841	17189	16117	15421	11836	9728	10707	11173	12405	8706	6010	3935	140068
HI	1505	1621	1636	1689	1591	1338	1425	1426	1384	1162	782	624	16183
ID	1863	1889	2038	1993	1574	1288	1413	1342	1167	1356	1002	767	17692
IL	24421	26771	27378	25390	21159	19085	20380	19808	19674	17521	14065	9339	244991
IN	9067	9374	9193	9304	7797	7009	6948	6958	7204	6475	4931	3453	87713
KY	5137	5118	4657	4506	3382	2602	2908	2746	3517	2862	2149	1465	41049
MA	12956	13238	11885	10482	7873	6545	6917	6864	8226	7182	6370	4427	102965
MD	11441	12331	11719	10593	8463	7106	7494	7225	9544	7216	4458	3003	100593
ME	521	598	650	599	497	417	474	466	437	473	441	386	5959
MI	9473	8987	8970	8673	7352	6491	6898	6786	7156	7088	5935	5477	89286
MN	8085	8552	8686	7954	6274	5412	5341	5140	5061	4427	3731	2839	71502
MO	4336	4505	4058	3877	3200	2647	2553	2404	2526	2129	1758	1307	35300
MP	73	117	116	122	120	153	148	134	149	87	87	48	1354
MT	248	275	328	384	335	268	291	265	234	162	171	102	3063
NC	16645	14756	14093	14106	12873	11517	11993	12077	14354	10881	7314	4323	144932
ND	451	480	506	490	383	385	284	324	336	296	210	136	4281
NH	553	630	519	516	406	319	368	360	346	330	315	209	4871
NJ	14291	14506	13341	12135	10019	8594	8639	8496	9169	8414	7352	5659	120615
NM	4259	4743	4769	4007	4050	4027	4322	4650	5208	4419	3431	2491	50376
NV	5833	6391	6186	6276	4794	4439	4858	4600	4232	4280	4122	3425	59436
OK	5412	6801	6764	6419	5601	4951	5296	5358	5622	4784	3782	2666	63456
PA	9503	9668	8848	8176	7240	6657	6769	6883	7415	6513	5516	4731	87919
PW	N/A	N/A	N/A	85	117	104	103	116	N/A	N/A	N/A	N/A	525
RI	1696	1694	1569	1691	1305	1250	1365	1410	1572	1548	1329	937	17366
SC	5304	5397	5293	5053	4154	4019	4143	4389	4949	4218	3260	2210	52389
SD	852	870	725	720	514	446	477	459	536	429	338	199	6565
TN	8685	8946	7560	6950	5839	4898	4959	4579	5424	4422	3628	2294	68184
UT	4856	5333	5477	5391	4602	4037	4977	4775	4708	4498	3535	2296	54485
VA	13243	15258	14188	13827	10477	8747	8998	8174	9906	8439	6927	4243	122427

State	G 1	G 2	G 3	G 4	G 5	G 6	G 7	G 8	G 9	G 10	G 11	G 12	Total
VI	90	69	104	108	114	97	108	92	139	87	69	42	1119
VT	161	170	154	179	122	129	125	120	133	122	108	88	1611
WA	13654	15054	15488	15566	12849	10735	10569	9645	9149	8378	7209	6264	134560
WI	5000	5389	5306	5499	4720	4343	4459	4415	4149	3753	3363	2531	52927
WY	260	269	281	266	203	169	179	173	181	160	167	153	2461
Total	235874	247413	237662	226083	186925	162872	169938	167095	180301	154277	122913	88406	2179759

Table 1.1.2.2

Participation by Grade by Gender, S602 Online

Grade	Statistic	Female	Male	Gender Missing	Total
1	Count	95582	104011	36281	235874
1	% within Grade	40.52%	44.10%	15.38%	100.00%
2	Count	100939	110252	36222	247413
2	% within Grade	40.80%	44.56%	14.64%	100.00%
3	Count	96569	105538	35555	237662
3	% within Grade	40.63%	44.41%	14.96%	100.00%
4	Count	89086	101337	35660	226083
4	% within Grade	39.40%	44.82%	15.77%	100.00%
5	Count	71442	84643	30840	186925
5	% within Grade	38.22%	45.28%	16.50%	100.00%
6	Count	61318	74710	26844	162872
6	% within Grade	37.65%	45.87%	16.48%	100.00%
7	Count	64601	78017	27320	169938
7	% within Grade	38.01%	45.91%	16.08%	100.00%
8	Count	63308	77548	26239	167095
8	% within Grade	37.89%	46.41%	15.70%	100.00%
9	Count	66959	84015	29327	180301
9	% within Grade	37.14%	46.60%	16.27%	100.00%
10	Count	58022	72084	24171	154277
10	% within Grade	37.61%	46.72%	15.67%	100.00%
11	Count	46331	57753	18829	122913
11	% within Grade	37.69%	46.99%	15.32%	100.00%
12	Count	34473	40707	13226	88406
12	% within Grade	38.99%	46.05%	14.96%	100.00%

Table 1.1.2.3**Participation by Grade by Ethnicity, S602 Online**

Grade	Statistic	Hispanic	Non-Hispanic	Unknown	Total
1	Count	152560	68152	15162	235874
1	% within Grade	64.68%	28.89%	6.43%	100.00%
2	Count	161386	70621	15406	247413
2	% within Grade	65.23%	28.54%	6.23%	100.00%
3	Count	155908	67279	14475	237662
3	% within Grade	65.60%	28.31%	6.09%	100.00%
4	Count	146964	62754	16365	226083
4	% within Grade	65.00%	27.76%	7.24%	100.00%
5	Count	124482	47975	14468	186925
5	% within Grade	66.59%	25.67%	7.74%	100.00%
6	Count	109300	39510	14062	162872
6	% within Grade	67.11%	24.26%	8.63%	100.00%
7	Count	115037	40506	14395	169938
7	% within Grade	67.69%	23.84%	8.47%	100.00%
8	Count	113645	39095	14355	167095
8	% within Grade	68.01%	23.40%	8.59%	100.00%
9	Count	122685	40014	17602	180301
9	% within Grade	68.04%	22.19%	9.76%	100.00%
10	Count	106287	35024	12966	154277
10	% within Grade	68.89%	22.70%	8.40%	100.00%
11	Count	83409	29107	10397	122913
11	% within Grade	67.86%	23.68%	8.46%	100.00%
12	Count	57728	22958	7720	88406
12	% within Grade	65.30%	25.97%	8.73%	100.00%

Table 1.1.2.4**Participation by Grade by Tier by Domain, S602 Online**

Grade	Tier	Writing	Speaking
01	PA	N/A	15259
01	A	209689	118617
01	BC	26157	101996
01	Total	235846	235872
02	PA	N/A	9641
02	A	87892	80831
02	BC	159485	156939
02	Total	247377	247411
03	PA	N/A	21084
03	A	73370	72488
03	BC	164271	144089
03	Total	237641	237661
04	PA	N/A	4147
04	A	59748	46837
04	BC	166333	175098
04	Total	226081	226082
05	PA	N/A	8815
05	A	56477	35201
05	BC	130445	142907
05	Total	186922	186923
06	PA	N/A	6313
06	A	64364	38138
06	BC	98505	118418
06	Total	162869	162869
07	PA	N/A	10496
07	A	76628	27965
07	BC	93306	131471
07	Total	169934	169932
08	PA	N/A	14936
08	A	77033	50673
08	BC	90055	101480
08	Total	167088	167089
09	PA	N/A	9369
09	A	74605	101101
09	BC	105690	69826
09	Total	180295	180296
10	PA	N/A	9529
10	A	56234	64419
10	BC	98038	80325

Grade	Tier	Writing	Speaking
10	Total	154272	154273
11	PA	N/A	11423
11	A	44811	26612
11	BC	78095	84873
11	Total	122906	122908
12	PA	N/A	8652
12	A	28930	41628
12	BC	59470	38119
12	Total	88400	88399

1.2 Scale Score Results

This section provides information on students' scale score results.

1.2.1 Mean Scale Score Across Domain and Composite Score by Cluster

This section shows mean (average) scale scores by grade-level cluster across the eight scores awarded, first for the four domains (Listening, Reading, Writing, and Speaking) and then for the four composites (Oral Language, Literacy, Comprehension, and Overall Composite). The mean scale scores are expected to increase as grade increases, as ACCESS is vertically scaled, but there is also an intersection between this principle and the population of test-takers.

In this section, under each average, the number of students in each group is also given. In Table 1.2.1.1, the order of average scale scores among single domains in descending order were Listening, Reading, Writing, and then Speaking except cluster 2–3. Cluster 4–5 showed the highest average scale score in Listening domain across all clusters.

Table 1.2.1.1

Mean Scale Scores by Cluster, S602 Online

Cluster	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
1	Mean	296.86	285.94	235.6	233.33	265.26	260.74	289.23	261.91
1	N	213555	223101	235734	214405	196145	223044	203660	187784
2-3	Mean	313.73	321.51	284.96	264.8	289.57	303.1	319.32	298.95
2-3	N	436928	452156	484822	445591	405599	452015	412025	384167
4-5	Mean	394.93	344.67	320.4	307.52	351.63	332.35	359.99	338.1
4-5	N	373317	374121	381687	377729	345642	351554	344035	304410
6-8	Mean	386.28	347.27	314.76	303.61	345.21	330.94	359.33	335.09
6-6	N	435520	446485	469133	448077	398577	427020	401324	358346
9-12	Mean	388.92	378.45	341.63	303.58	346.44	360.11	381.85	355.76
9-12	N	493338	489225	511940	501163	458146	465480	451351	405156

Table 1.2.1.2 demonstrates that groups made up of female students performed better than groups of male students in clusters 1 and 2–3.

Table 1.2.1.2

Mean Scale Scores by Gender, S602 Online

Cluster	Gender	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
1	F	Mean	299.83	286.29	239.59	238.32	269.14	262.84	290.31	264.41
1	F	N	87412	90316	95509	87949	81131	90284	83281	77586
1	M	Mean	293.83	286	232.36	229.32	261.85	259.19	288.42	259.89
1	M	N	93807	98622	103953	94140	85772	98604	89679	82294
1	Missing	Mean	297.63	284.89	234.36	231.4	264.51	259.67	288.69	260.93
1	Missing	N	32336	34163	36272	32316	29242	34156	30700	27904
2-3	F	Mean	314.56	322.23	290.1	270.01	292.5	305.98	320.03	301.7
2-3	F	N	179545	183989	197417	182966	167823	183949	169124	158824
2-3	M	Mean	313.07	321.39	281.21	261.1	287.44	301.24	319.07	297.06
2-3	M	N	194115	202353	215656	197955	179940	202277	184004	171176
2-3	Missing	Mean	313.43	319.9	282.09	261.35	287.71	300.76	318.06	296.89
2-3	Missing	N	63268	65814	71749	64670	57836	65789	58897	54167
4-5	F	Mean	393.49	345.22	325.51	310	352.08	335.19	359.9	340.06
4-5	F	N	146397	145735	148358	147878	136259	136947	135087	120022
4-5	M	Mean	395.36	344.03	316.9	305.85	351.09	330.32	359.71	336.66
4-5	M	N	168273	170041	172767	170283	155905	160174	156301	138438
4-5	Missing	Mean	397.3	345.16	317.87	306.13	352.02	331.16	361.04	337.3
4-5	Missing	N	58647	58345	60562	59568	53478	54433	52647	45950
6-8	F	Mean	384.7	348.61	319.09	304.32	344.66	333.79	359.72	336.74
6-8	F	N	166815	169194	177575	169991	152581	161795	153467	136976
6-8	M	Mean	387.48	346.43	312.17	304.05	346.13	329.24	359.12	334.25
6-8	M	N	199923	206977	216745	207305	183477	197975	184935	165241
6-8	Missing	Mean	386.61	346.53	312	300.61	343.87	329.08	358.99	333.53
6-8	Missing	N	68782	70314	74813	70781	62519	67250	62922	56129
9-12	F	Mean	387.42	380.18	345.1	305.87	346.75	362.75	382.59	357.59
9-12	F	N	186974	183706	192544	188863	173456	174665	170401	152737
9-12	M	Mean	389.76	377.04	339.69	302.43	346.33	358.41	381.09	354.57
9-12	M	N	228534	228597	238995	233818	212447	217456	209631	188048
9-12	Missing	Mean	390.06	378.53	339.05	301.52	346.04	358.85	382.31	354.91
9-12	Missing	N	77830	76922	80401	78482	72243	73359	71319	64371

Table 1.2.1.3 presents scale score performance by ethnic groups. The top three performing ethnic groups were Asian students, White students, and multiracial students in most domains and clusters. Tables 1.2.1.4 through 1.2.1.7 show this information by gender, and by race and ethnicity.

Table 1.2.1.3

Mean Scale Scores by Ethnicity for Cluster 1, S602 Online

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	291.16	281.91	228.2	227.5	259.48	255.06	284.67	256.16
Hispanic (of any Race)	N	138077	144663	152473	139162	127202	144618	131939	121948
Non-Hispanic American Indian	Mean	299.56	285.29	234.49	234.89	266.33	260.03	289.42	261.41
Non-Hispanic American Indian	N	1482	1514	1632	1448	1324	1512	1388	1254
Non-Hispanic Asian	Mean	316.88	302.9	263.58	251.37	284.43	283.34	307.19	283.64
Non-Hispanic Asian	N	26117	26991	28536	25965	23989	26988	24905	23000
Non-Hispanic Black	Mean	301.48	290.1	242.19	250.46	276.08	266.17	293.59	269.12
Non-Hispanic Black	N	11250	11765	12482	11200	10220	11763	10700	9744
Non-Hispanic Multiracial	Mean	312.37	293.81	247.86	248.65	281.24	270.83	299.92	274.52
Non-Hispanic Multiracial	N	1054	1114	1168	1079	986	1114	1015	953
Non-Hispanic Pacific Islander	Mean	286.43	282.93	237.35	231.14	258.96	260.06	283.63	259.12
Non-Hispanic Pacific Islander	N	1552	1604	1726	1543	1406	1604	1457	1329
Non-Hispanic White	Mean	312.61	291.46	251.24	245.53	279.43	271.21	297.87	273.55
Non-Hispanic White	N	20403	21120	22568	20393	18630	21117	19258	17672
Unknown	Mean	288.52	282.99	227.6	224.95	256.65	255.08	284.46	255.01
Unknown	N	13620	14330	15149	13615	12388	14328	12998	11884

Table 1.2.1.4**Mean Scale Scores by Ethnicity for Cluster 2-3, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	308.39	318.32	279.29	260.06	284.5	298.66	315.47	294.29
Hispanic (of any Race)	N	286529	296491	317143	292220	266388	296394	270557	252607
Non-Hispanic American Indian	Mean	317.02	318.18	284.14	265.23	291.34	300.96	317.85	297.57
Non-Hispanic American Indian	N	3230	3388	3679	3332	2979	3386	3020	2797
Non-Hispanic Asian	Mean	333.63	335.4	307.82	281.26	307.86	321.76	335.08	317.68
Non-Hispanic Asian	N	50870	52195	55854	51497	47391	52184	48107	44996
Non-Hispanic Black	Mean	319.13	324.05	290.67	278.5	299.39	307.28	322.84	305.14
Non-Hispanic Black	N	22837	23804	25629	23264	20996	23799	21517	19877
Non-Hispanic Multiracial	Mean	331.68	329.32	297.67	281.41	306.95	313.66	330.44	311.86
Non-Hispanic Multiracial	N	2127	2203	2338	2170	1987	2202	2025	1895
Non-Hispanic Pacific Islander	Mean	303.62	316.66	292.16	258.24	281.34	304.34	313.11	297.45
Non-Hispanic Pacific Islander	N	3465	3580	3852	3536	3209	3579	3253	3034
Non-Hispanic White	Mean	328.73	328.12	298.47	277.34	303.32	313.16	328.42	310.05
Non-Hispanic White	N	41342	42597	46471	42254	38121	42584	38447	35651
Unknown	Mean	304.78	317.62	274.69	252.78	278.98	295.74	313.73	290.41
Unknown	N	26528	27898	29856	27318	24528	27887	25099	23310

Table 1.2.1.5**Mean Scale Scores by Ethnicity for Cluster 4-5, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	391.99	342.53	317.7	304.65	348.71	329.95	357.61	335.56
Hispanic (of any Race)	N	245939	246558	251043	248910	228138	231664	227162	201311
Non-Hispanic American Indian	Mean	396.95	340.34	317.12	303.65	350.6	328.26	357.57	334.99
Non-Hispanic American Indian	N	3335	3307	3456	3362	3021	3102	2994	2588
Non-Hispanic Asian	Mean	412.39	358.85	340.99	323.8	368.57	350	375.16	355.61
Non-Hispanic Asian	N	38735	38618	39041	38965	36104	36244	35875	31898
Non-Hispanic Black	Mean	401.94	346.26	322.42	321.96	362.58	334.22	363.44	343.17
Non-Hispanic Black	N	19733	19753	20178	19926	18105	18415	18028	15704
Non-Hispanic Multiracial	Mean	407.95	351.25	328.92	320.7	364.26	339.6	368.62	346.66
Non-Hispanic Multiracial	N	1521	1531	1570	1552	1432	1458	1416	1296
Non-Hispanic Pacific Islander	Mean	393.83	342.85	326.61	305.06	349.51	334.76	358.11	339.05
Non-Hispanic Pacific Islander	N	3615	3588	3670	3665	3325	3332	3266	2868
Non-Hispanic White	Mean	406.21	351.4	330.65	320.67	363.95	340.83	368.11	347.62
Non-Hispanic White	N	32925	32603	33116	33341	30204	30085	29736	25728
Unknown	Mean	377.29	335.49	302.51	284.51	330.91	318.39	348.03	321.69
Unknown	N	27514	28163	29613	28008	25313	27254	25558	23017

Table 1.2.1.6**Mean Scale Scores by Ethnicity for Cluster 6-8, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	384.57	345.86	313.84	301.02	343.07	329.8	357.79	333.65
Hispanic (of any Race)	N	295292	303218	318245	304401	271298	290663	272924	244837
Non-Hispanic American Indian	Mean	392.93	345.95	315.96	304.21	349.02	330.81	360.57	336.4
Non-Hispanic American Indian	N	4112	4307	4564	4263	3669	4078	3768	3269
Non-Hispanic Asian	Mean	400.94	360.95	331.64	325.4	363.61	346.38	373.39	351.62
Non-Hispanic Asian	N	36004	36308	38125	36335	32756	34730	33022	29384
Non-Hispanic Black	Mean	393.11	349.88	315.37	315.49	354.49	332.55	363.28	339.15
Non-Hispanic Black	N	22590	23189	24392	23245	20448	22007	20654	18190
Non-Hispanic Multiracial	Mean	396.81	352.58	320.86	316.68	357.08	337.15	366.53	343.22
Non-Hispanic Multiracial	N	1641	1646	1719	1639	1500	1578	1521	1353
Non-Hispanic Pacific Islander	Mean	389	347.35	321.29	306.95	348.55	334.27	360.69	338.82
Non-Hispanic Pacific Islander	N	4358	4471	4854	4682	3897	4123	3809	3284
Non-Hispanic White	Mean	394.99	352.94	322.42	317.04	356.43	337.69	366.12	343.27
Non-Hispanic White	N	34350	34897	36806	35099	30915	32922	31173	27136
Unknown	Mean	371.94	338.71	297.63	283	327.51	317.8	348.91	320.3
Unknown	N	37173	38449	40428	38413	34094	36919	34453	30893

Table 1.2.1.7**Mean Scale Scores by Ethnicity for Cluster 9-12, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	386.07	376.64	340.81	299.95	343.2	358.78	379.69	353.84
Hispanic (of any Race)	N	335311	333390	348207	341144	312339	317915	308185	277761
Non-Hispanic American Indian	Mean	397.69	381.78	348.89	308.89	353.86	365.52	386.82	361.61
Non-Hispanic American Indian	N	4870	5155	5390	5129	4475	4916	4487	3995
Non-Hispanic Asian	Mean	407.38	392.75	358.63	329.35	368.62	375.87	397.46	373.48
Non-Hispanic Asian	N	36657	35793	37267	36495	33762	33795	33343	29550
Non-Hispanic Black	Mean	395.53	382.53	342.44	317.79	356.8	362.66	386.74	360.75
Non-Hispanic Black	N	28993	28492	30019	29677	26935	26909	26132	23338
Non-Hispanic Multiracial	Mean	401.39	385.48	347.54	318.13	360.14	366.97	390.75	364.93
Non-Hispanic Multiracial	N	1685	1676	1729	1707	1577	1597	1552	1395
Non-Hispanic Pacific Islander	Mean	393.35	377.66	351.29	305.72	349.62	364.63	382.6	359.78
Non-Hispanic Pacific Islander	N	4262	4198	4392	4242	3781	3883	3785	3189
Non-Hispanic White	Mean	400.34	384.9	345.76	315.78	358.29	365.64	390.11	363.6
Non-Hispanic White	N	38173	37005	39478	38248	35044	35237	34202	30440
Unknown	Mean	378.99	371.84	327.82	288.97	334.05	349.67	374.18	344.63
Unknown	N	43387	43516	45458	44521	40233	41228	39665	35488

1.2.2 Mean Scale Score Across Domain and Composite Score by Grade

This section provides parallel information to the prior section, with mean scale scores broken down by grade rather than by grade-level cluster. Table 1.2.2.1 shows the increment of scale scores by grade, which peaked at grade 5 in the Listening domain and at grade 12 for all other domains. Table 1.2.2.2 demonstrates student performance by grade and gender. Tables 1.2.2.3 through 1.2.2.14 show student performance by race and ethnicity.

Table 1.2.2.1**Mean Scale Scores by Grade, S602 Online**

Grade	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
1	Mean	296.86	285.94	235.6	233.33	265.26	260.74	289.23	261.91
1	N	213555	223101	235734	214405	196145	223044	203660	187784
2	Mean	303.79	316.6	274.91	257.38	280.88	295.56	312.79	291.01
2	N	221182	230737	247257	225607	204084	230649	208640	193331
3	Mean	323.92	326.63	295.42	272.4	298.38	310.96	326.01	307
3	N	215746	221419	237565	219984	201515	221366	203385	190836
4	Mean	393	343.29	316.73	308.16	351.04	329.76	358.44	336.15
4	N	203631	204270	208039	206145	188065	191245	187289	164672
5	Mean	397.25	346.34	324.79	306.75	352.32	335.43	361.83	340.39
5	N	169686	169851	173648	171584	157577	160309	156746	139738
6	Mean	379.43	339.95	305.96	299.74	340.03	322.91	352.18	328.07
6	N	140833	145493	152901	145767	128760	139105	129782	115623
7	Mean	387.12	347.41	315.29	303.72	345.7	331.28	359.65	335.48
7	N	147693	151390	159502	151322	134553	144880	135867	121006
8	Mean	392	354.25	322.81	307.23	349.65	338.41	365.84	341.36
8	N	146994	149602	156730	150988	135264	143035	135675	121717
9	Mean	382.8	373.39	335.05	296.58	339.84	354.16	376.4	349.6
9	N	161437	160871	168622	165462	149793	152812	147401	132080
10	Mean	389.28	379.01	341.82	304.48	347.07	360.48	382.33	356.18
10	N	139476	138099	144550	141342	129330	131267	127547	114235
11	Mean	392.68	381.78	346.62	307.58	350.34	364.34	385.36	359.91
11	N	111575	110332	115320	112060	102991	105019	102031	91135
12	Mean	395.31	383.08	347.69	310.68	353.1	365.53	387.02	361.48
12	N	80850	79923	83448	82299	76032	76382	74372	67706

Table 1.2.2.2**Mean Scale Scores by Grade by Gender, S602 Online**

Grade	Gender	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
1	F	Mean	299.83	286.29	239.59	238.32	269.14	262.84	290.31	264.41
1	F	N	87412	90316	95509	87949	81131	90284	83281	77586
1	M	Mean	293.83	286	232.36	229.32	261.85	259.19	288.42	259.89
1	M	N	93807	98622	103953	94140	85772	98604	89679	82294
1	Missing	Mean	297.63	284.89	234.36	231.4	264.51	259.67	288.69	260.93
1	Missing	N	32336	34163	36272	32316	29242	34156	30700	27904
2	F	Mean	305.15	317.19	279.79	262.53	284.03	298.22	313.56	293.64
2	F	N	91038	94002	100883	92927	84669	93980	85701	80077
2	M	Mean	302.89	316.68	271.43	253.75	278.67	293.93	312.6	289.26
2	M	N	98382	103333	110173	100368	90628	103286	93222	86150
2	Missing	Mean	302.69	314.72	271.91	253.88	278.55	293.1	311.15	288.79
2	Missing	N	31762	33402	36201	32312	28787	33383	29717	27104
3	F	Mean	324.23	327.49	300.87	277.74	301.12	314.09	326.67	309.89
3	F	N	88507	89987	96534	90039	83154	89969	83423	78747
3	M	Mean	323.53	326.3	291.43	268.66	296.34	308.87	325.71	304.96
3	M	N	95733	99020	105483	97587	89312	98991	90782	85026
3	Missing	Mean	324.26	325.24	292.46	268.81	296.78	308.64	325.09	305
3	Missing	N	31506	32412	35548	32358	29049	32406	29180	27063
4	F	Mean	391.94	343.81	322.04	311.63	352.15	332.71	358.47	338.37
4	F	N	81081	80556	81883	81834	75285	75393	74596	65820
4	M	Mean	393.43	342.9	313.2	305.85	350.23	327.87	358.35	334.76
4	M	N	91371	92596	93812	92443	84391	86981	84874	74742
4	Missing	Mean	394.53	343.09	313.55	305.91	350.54	327.77	358.64	334.37
4	Missing	N	31179	31118	32344	31868	28389	28871	27819	24110
5	F	Mean	395.41	346.98	329.78	307.98	351.99	338.22	361.66	342.12
5	F	N	65316	65179	66475	66044	60974	61554	60491	54202
5	M	Mean	397.65	345.39	321.3	305.84	352.12	333.24	361.32	338.87
5	M	N	76902	77445	78955	77840	71514	73193	71427	63696
5	Missing	Mean	400.46	347.53	322.83	306.39	353.7	334.98	363.73	340.54
5	Missing	N	27468	27227	28218	27700	25089	25562	24828	21840
6	F	Mean	377.42	340.75	310.03	299.91	339.01	325.34	352.03	329.3
6	F	N	53854	54910	57603	55233	49371	52548	49590	44279
6	M	Mean	380.58	339.19	303.14	300.18	340.94	321.16	352.03	327.21
6	M	N	64489	67338	70331	67108	59071	64337	59766	53240
6	Missing	Mean	380.94	340.26	304.52	298.06	339.88	322.21	352.95	327.62
6	Missing	N	22490	23245	24967	23426	20318	22220	20426	18104
7	F	Mean	385.6	349	319.78	304.43	345.25	334.35	360.25	337.34
7	F	N	56776	57579	60633	57703	51733	55093	52127	46439

Grade	Gender	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
7	M	Mean	388.36	346.41	312.58	304.3	346.68	329.44	359.36	334.56
7	M	N	67493	69980	73490	69801	61672	67031	62353	55591
7	Missing	Mean	387.23	346.52	312.41	300.3	343.98	329.23	359.07	333.61
7	Missing	N	23424	23831	25379	23818	21148	22756	21387	18976
8	F	Mean	390.77	355.84	327.18	308.47	349.49	341.43	366.54	343.26
8	F	N	56185	56705	59339	57055	51477	54154	51750	46258
8	M	Mean	393.17	353.44	320.47	307.48	350.47	336.85	365.64	340.58
8	M	N	67941	69659	72924	70396	62734	66607	62816	56410
8	Missing	Mean	391.55	352.82	319.22	303.44	347.61	335.78	364.75	339.06
8	Missing	N	22868	23238	24467	23537	21053	22274	21109	19049
9	F	Mean	381.3	374.91	338.76	298.37	339.9	356.86	377.04	351.45
9	F	N	60249	59517	62452	61506	55897	56461	54831	49082
9	M	Mean	383.69	372.19	333.26	296.25	340.13	352.61	375.78	348.61
9	M	N	74701	75033	78652	77017	69284	71279	68313	61109
9	Missing	Mean	383.73	373.37	331.76	293.44	338.9	352.46	376.69	348.25
9	Missing	N	26487	26321	27518	26939	24612	25072	24257	21889
10	F	Mean	387.66	380.52	345.13	306.65	347.19	362.93	382.87	357.76
10	F	N	52748	51712	54276	53134	48872	49128	48036	42976
10	M	Mean	389.84	377.5	339.67	302.99	346.67	358.62	381.4	354.78
10	M	N	64668	64633	67575	66128	60084	61382	59282	53078
10	Missing	Mean	391.5	379.92	340.31	303.72	347.94	360.19	383.78	356.51
10	Missing	N	22060	21754	22699	22080	20374	20757	20229	18181
11	F	Mean	390.73	383.42	349.51	309.76	350.43	366.62	385.85	361.38
11	F	N	42308	41423	43361	42147	38962	39428	38546	34367
11	M	Mean	393.82	380.35	344.93	306.08	350.22	362.77	384.73	358.84
11	M	N	52066	52011	54254	52706	48102	49480	47786	42617
11	Missing	Mean	394.04	382.18	344.73	306.81	350.5	363.6	386.07	359.58
11	Missing	N	17201	16898	17705	17207	15927	16111	15699	14151
12	F	Mean	394.23	385.38	351.39	313.84	354.08	368.52	388.33	363.78
12	F	N	31669	31054	32455	32076	29725	29648	28988	26312
12	M	Mean	396.14	381.43	345.51	308.9	352.68	363.61	386.06	360.04
12	M	N	37099	36920	38514	37967	34977	35315	34250	31244
12	Missing	Mean	395.61	382.2	344.81	307.91	351.84	363.72	386.56	359.93
12	Missing	N	12082	11949	12479	12256	11330	11419	11134	10150

Table 1.2.2.3**Mean Scale Scores by Ethnicity for Grade 1, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	291.16	281.91	228.2	227.5	259.48	255.06	284.67	256.16
Hispanic (of any Race)	N	138077	144663	152473	139162	127202	144618	131939	121948
Non-Hispanic American Indian	Mean	299.56	285.29	234.49	234.89	266.33	260.03	289.42	261.41
Non-Hispanic American Indian	N	1482	1514	1632	1448	1324	1512	1388	1254
Non-Hispanic Asian	Mean	316.88	302.9	263.58	251.37	284.43	283.34	307.19	283.64
Non-Hispanic Asian	N	26117	26991	28536	25965	23989	26988	24905	23000
Non-Hispanic Black	Mean	301.48	290.1	242.19	250.46	276.08	266.17	293.59	269.12
Non-Hispanic Black	N	11250	11765	12482	11200	10220	11763	10700	9744
Non-Hispanic Multiracial	Mean	312.37	293.81	247.86	248.65	281.24	270.83	299.92	274.52
Non-Hispanic Multiracial	N	1054	1114	1168	1079	986	1114	1015	953
Non-Hispanic Pacific Islander	Mean	286.43	282.93	237.35	231.14	258.96	260.06	283.63	259.12
Non-Hispanic Pacific Islander	N	1552	1604	1726	1543	1406	1604	1457	1329
Non-Hispanic White	Mean	312.61	291.46	251.24	245.53	279.43	271.21	297.87	273.55
Non-Hispanic White	N	20403	21120	22568	20393	18630	21117	19258	17672
Unknown	Mean	288.52	282.99	227.6	224.95	256.65	255.08	284.46	255.01
Unknown	N	13620	14330	15149	13615	12388	14328	12998	11884

Table 1.2.2.4**Mean Scale Scores by Ethnicity for Grade 2, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	298.17	313.71	268.47	252.18	275.42	290.9	309.07	286.09
Hispanic (of any Race)	N	144692	150951	161296	147520	133666	150896	136751	126858
Non-Hispanic American Indian	Mean	307.99	315.07	275.21	259.43	283.91	294.84	312.87	291
Non-Hispanic American Indian	N	1632	1724	1869	1684	1500	1723	1527	1411
Non-Hispanic Asian	Mean	323.89	328.65	299.84	274.51	299.61	314.34	327.35	309.97
Non-Hispanic Asian	N	26207	27041	28932	26546	24310	27031	24778	23050
Non-Hispanic Black	Mean	309.25	318.93	281.49	272.18	291.31	300.05	316.07	297.48
Non-Hispanic Black	N	11434	12045	12927	11602	10410	12043	10792	9868
Non-Hispanic Multiracial	Mean	322.69	323.23	288.6	274.92	299.46	305.99	323.4	304.07
Non-Hispanic Multiracial	N	1078	1129	1192	1114	1013	1128	1029	968
Non-Hispanic Pacific Islander	Mean	293.21	312.51	281.34	251.36	272.44	296.49	306.94	288.89
Non-Hispanic Pacific Islander	N	1730	1799	1955	1773	1583	1798	1610	1485
Non-Hispanic White	Mean	318.97	322.19	289.46	269.9	294.69	305.58	321.21	302.1
Non-Hispanic White	N	20833	21651	23698	21409	19139	21641	19310	17837
Unknown	Mean	296.34	314.15	265.74	247.37	272.1	289.45	308.68	283.94
Unknown	N	13576	14397	15388	13959	12463	14389	12843	11854

Table 1.2.2.5**Mean Scale Scores by Ethnicity for Grade 3, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	318.81	323.1	290.48	268.1	293.65	306.71	322	302.57
Hispanic (of any Race)	N	141837	145540	155847	144700	132722	145498	133806	125749
Non-Hispanic American Indian	Mean	326.24	321.4	293.37	271.16	298.87	307.3	322.95	304.27
Non-Hispanic American Indian	N	1598	1664	1810	1648	1479	1663	1493	1386
Non-Hispanic Asian	Mean	343.98	342.66	316.39	288.43	316.55	329.73	343.29	325.78
Non-Hispanic Asian	N	24663	25154	26922	24951	23081	25153	23329	21946
Non-Hispanic Black	Mean	329.04	329.3	300.02	284.78	307.34	314.69	329.65	312.69
Non-Hispanic Black	N	11403	11759	12702	11662	10586	11756	10725	10009
Non-Hispanic Multiracial	Mean	340.91	335.72	307.11	288.24	314.75	321.72	337.72	320
Non-Hispanic Multiracial	N	1049	1074	1146	1056	974	1074	996	927
Non-Hispanic Pacific Islander	Mean	313.99	320.84	303.31	265.16	290.01	312.26	319.16	305.66
Non-Hispanic Pacific Islander	N	1735	1781	1897	1763	1626	1781	1643	1549
Non-Hispanic White	Mean	338.64	334.25	307.84	284.98	312.02	320.99	335.69	318.02
Non-Hispanic White	N	20509	20946	22773	20845	18982	20943	19137	17814
Unknown	Mean	313.63	321.33	284.22	258.43	286.09	302.45	319.02	297.11
Unknown	N	12952	13501	14468	13359	12065	13498	12256	11456

Table 1.2.2.6**Mean Scale Scores by Ethnicity for Grade 4, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	389.58	340.88	313.49	305.04	347.74	326.96	355.71	333.19
Hispanic (of any Race)	N	132662	133118	135371	134378	122767	124637	122261	107656
Non-Hispanic American Indian	Mean	390.42	337.51	310.33	301.53	346.76	323.64	353.81	330.75
Non-Hispanic American Indian	N	1715	1700	1760	1741	1565	1586	1529	1321
Non-Hispanic Asian	Mean	411.15	357.58	338.18	324.25	368.15	347.84	373.87	353.94
Non-Hispanic Asian	N	22485	22408	22577	22535	20874	20940	20807	18346
Non-Hispanic Black	Mean	400.3	345.09	319.2	322.37	362.02	332	362.18	341.49
Non-Hispanic Black	N	10858	10925	11110	10986	9930	10155	9936	8605
Non-Hispanic Multiracial	Mean	406.74	349.98	325.1	321.83	364.52	337.05	367.56	345.28
Non-Hispanic Multiracial	N	836	841	862	849	784	799	779	709
Non-Hispanic Pacific Islander	Mean	387.95	338.97	318.69	300.68	344.47	328.71	353.66	333.28
Non-Hispanic Pacific Islander	N	1826	1855	1880	1860	1674	1712	1669	1455
Non-Hispanic White	Mean	404.51	350.09	327.8	321.45	363.63	338.65	366.63	346
Non-Hispanic White	N	18797	18499	18813	18990	17210	17004	16890	14539
Unknown	Mean	375.9	334.37	298.83	285.29	330.68	316.04	346.95	320.16
Unknown	N	14452	14924	15666	14806	13261	14412	13418	12041

Table 1.2.2.7**Mean Scale Scores by Ethnicity for Grade 5, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	394.82	344.47	322.62	304.19	349.83	333.43	359.81	338.27
Hispanic (of any Race)	N	113277	113440	115672	114532	105371	107027	104901	93655
Non-Hispanic American Indian	Mean	403.86	343.32	324.17	305.94	354.72	333.09	361.5	339.4
Non-Hispanic American Indian	N	1620	1607	1696	1621	1456	1516	1465	1267
Non-Hispanic Asian	Mean	414.11	360.61	344.83	323.19	369.16	352.94	376.95	357.88
Non-Hispanic Asian	N	16250	16210	16464	16430	15230	15304	15068	13552
Non-Hispanic Black	Mean	403.95	347.71	326.36	321.45	363.28	336.96	364.98	345.21
Non-Hispanic Black	N	8875	8828	9068	8940	8175	8260	8092	7099
Non-Hispanic Multiracial	Mean	409.42	352.81	333.58	319.35	363.95	342.69	369.92	348.32
Non-Hispanic Multiracial	N	685	690	708	703	648	659	637	587
Non-Hispanic Pacific Islander	Mean	399.83	347.01	334.93	309.57	354.62	341.15	362.76	344.99
Non-Hispanic Pacific Islander	N	1789	1733	1790	1805	1651	1620	1597	1413
Non-Hispanic White	Mean	408.47	353.11	334.41	319.63	364.37	343.65	370.05	349.71
Non-Hispanic White	N	14128	14104	14303	14351	12994	13081	12846	11189
Unknown	Mean	378.83	336.74	306.63	283.63	331.16	321.04	349.24	323.38
Unknown	N	13062	13239	13947	13202	12052	12842	12140	10976

Table 1.2.2.8**Mean Scale Scores by Ethnicity for Grade 6, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	378.02	338.58	305.1	297.62	338.29	321.81	350.76	326.76
Hispanic (of any Race)	N	94716	97894	102852	98305	86969	93768	87470	78322
Non-Hispanic American Indian	Mean	385.19	337.71	308.2	302.54	343.95	322.82	352.59	329.45
Non-Hispanic American Indian	N	1355	1431	1506	1408	1199	1351	1237	1069
Non-Hispanic Asian	Mean	391.22	352.1	321.24	316.9	354.66	336.79	364.33	342.43
Non-Hispanic Asian	N	12219	12418	13012	12382	11128	11890	11253	10008
Non-Hispanic Black	Mean	384.69	341.96	305.62	310.33	347.85	323.71	355.23	331.01
Non-Hispanic Black	N	7140	7420	7808	7370	6435	7054	6552	5739
Non-Hispanic Multiracial	Mean	389.95	345.09	312.95	313.06	351.34	329.68	359.26	336.11
Non-Hispanic Multiracial	N	553	561	577	549	495	532	514	445
Non-Hispanic Pacific Islander	Mean	383.12	341.73	313.45	304.93	344.74	327.8	355.07	333.39
Non-Hispanic Pacific Islander	N	1467	1536	1645	1584	1315	1412	1302	1113
Non-Hispanic White	Mean	386.95	345.06	313.23	312.01	349.91	329.19	357.98	335.37
Non-Hispanic White	N	11265	11496	12187	11590	10136	10852	10146	8817
Unknown	Mean	366.89	332.67	289.76	280.34	324.01	310.96	343.32	314.87
Unknown	N	12118	12737	13314	12579	11083	12246	11308	10110

Table 1.2.2.9**Mean Scale Scores by Ethnicity for Grade 7, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	385.47	345.98	314.31	301.19	343.61	330.1	358.11	334.02
Hispanic (of any Race)	N	100187	102942	108304	102889	91575	98697	92450	82621
Non-Hispanic American Indian	Mean	392.6	344.76	315.62	301.96	348.05	330.03	359.27	335.64
Non-Hispanic American Indian	N	1325	1409	1499	1376	1175	1336	1216	1055
Non-Hispanic Asian	Mean	401.77	361.43	332.28	325.42	364.02	346.93	373.96	352.18
Non-Hispanic Asian	N	12117	12225	12898	12209	10983	11721	11069	9847
Non-Hispanic Black	Mean	393.67	349.53	315.57	315.66	354.93	332.5	363.22	339.41
Non-Hispanic Black	N	7750	7916	8370	7928	6998	7517	7050	6212
Non-Hispanic Multiracial	Mean	396.78	352.41	321.69	315.67	356.99	337.3	366.35	343.15
Non-Hispanic Multiracial	N	548	553	584	557	508	530	507	454
Non-Hispanic Pacific Islander	Mean	388.2	346.88	322.22	305.26	347.39	334.44	359.94	338.29
Non-Hispanic Pacific Islander	N	1488	1516	1646	1556	1311	1399	1296	1119
Non-Hispanic White	Mean	396.04	353.33	323.52	317.35	357.21	338.38	366.82	344.02
Non-Hispanic White	N	11736	11956	12623	11926	10509	11328	10689	9303
Unknown	Mean	372.63	338.89	297.99	282.9	327.76	318.12	349.26	320.6
Unknown	N	12542	12873	13578	12881	11494	12352	11590	10395

Table 1.2.2.10**Mean Scale Scores by Ethnicity for Grade 8, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	389.85	352.69	321.75	304.09	347.02	337.14	364.08	339.71
Hispanic (of any Race)	N	100389	102382	107089	103207	92754	98198	93004	83894
Non-Hispanic American Indian	Mean	400.55	355.12	323.79	307.89	354.59	339.33	369.29	343.57
Non-Hispanic American Indian	N	1432	1467	1559	1479	1295	1391	1315	1145
Non-Hispanic Asian	Mean	410.26	369.89	342.04	334.33	372.54	356.04	382.33	360.68
Non-Hispanic Asian	N	11668	11665	12215	11744	10645	11119	10700	9529
Non-Hispanic Black	Mean	400.34	357.73	324.44	320.11	360.15	340.98	370.81	346.39
Non-Hispanic Black	N	7700	7853	8214	7947	7015	7436	7052	6239
Non-Hispanic Multiracial	Mean	403.86	360.64	328.18	321.46	362.9	344.69	374.18	350.26
Non-Hispanic Multiracial	N	540	532	558	533	497	516	500	454
Non-Hispanic Pacific Islander	Mean	395.99	353.92	328.58	310.73	353.69	341.06	367.53	345.12
Non-Hispanic Pacific Islander	N	1403	1419	1563	1542	1271	1312	1211	1052
Non-Hispanic White	Mean	401.88	360.46	330.61	321.76	362.06	345.55	373.38	350.22
Non-Hispanic White	N	11349	11445	11996	11583	10270	10742	10338	9016
Unknown	Mean	376.15	344.53	305.02	285.7	330.64	324.28	354.03	325.27
Unknown	N	12513	12839	13536	12953	11517	12321	11555	10388

Table 1.2.2.11**Mean Scale Scores by Ethnicity for Grade 9, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	380.52	371.97	334.55	293.76	337.33	353.19	374.7	348.18
Hispanic (of any Race)	N	110178	109895	115103	113023	102539	104625	100966	90812
Non-Hispanic American Indian	Mean	388.41	374.99	342.28	300.73	344.54	358.62	378.89	353.72
Non-Hispanic American Indian	N	1543	1615	1688	1639	1442	1533	1424	1279
Non-Hispanic Asian	Mean	404.27	389.24	354.93	324.47	364.69	372.21	394.1	369.85
Non-Hispanic Asian	N	11435	11250	11745	11490	10483	10615	10375	9165
Non-Hispanic Black	Mean	388.64	376.23	336.18	309.83	349.19	356.35	380.2	353.99
Non-Hispanic Black	N	8671	8595	9044	8953	8040	8103	7813	6950
Non-Hispanic Multiracial	Mean	396.78	381.7	344.84	317.41	357.12	363.33	386.57	361.12
Non-Hispanic Multiracial	N	533	527	543	546	499	493	484	426
Non-Hispanic Pacific Islander	Mean	390.08	374.48	348.83	302.81	346.74	361.9	379.29	357.17
Non-Hispanic Pacific Islander	N	1503	1477	1566	1527	1344	1368	1322	1126
Non-Hispanic White	Mean	393.98	379.35	340.82	310.19	352.09	360.37	384.24	358
Non-Hispanic White	N	12027	11687	12510	12177	11041	11097	10719	9540
Unknown	Mean	369.56	365.47	316.95	277.11	323.29	340.87	366.73	335.07
Unknown	N	15547	15825	16423	16107	14405	14978	14298	12782

Table 1.2.2.12**Mean Scale Scores by Ethnicity for Grade 10, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	386.41	377.27	340.9	300.92	343.84	359.15	380.22	354.26
Hispanic (of any Race)	N	96258	95637	99837	97741	89523	91067	88442	79488
Non-Hispanic American Indian	Mean	398.16	381.81	347.48	304.4	352.2	364.65	387.29	360.85
Non-Hispanic American Indian	N	1343	1412	1477	1412	1232	1344	1234	1101
Non-Hispanic Asian	Mean	407	392.59	357.81	330.01	368.72	375.37	397.16	373.02
Non-Hispanic Asian	N	10052	9765	10196	9996	9301	9224	9138	8118
Non-Hispanic Black	Mean	395.97	383.04	342.7	318.08	357.16	363.08	387.17	360.93
Non-Hispanic Black	N	7985	7854	8294	8175	7391	7416	7198	6415
Non-Hispanic Multiracial	Mean	400.09	384.33	345.27	313.57	357.17	364.97	389.53	362.76
Non-Hispanic Multiracial	N	490	493	507	497	459	473	456	414
Non-Hispanic Pacific Islander	Mean	396.09	378.28	350.8	303.75	350.32	364.51	384.25	360.16
Non-Hispanic Pacific Islander	N	1205	1175	1228	1195	1071	1086	1067	898
Non-Hispanic White	Mean	400.86	385.58	346.3	317.36	359.39	366.25	390.73	364.4
Non-Hispanic White	N	10549	10208	10876	10510	9630	9683	9437	8314
Unknown	Mean	380.39	372.91	329.64	291.2	335.81	351.12	375.28	346.18
Unknown	N	11594	11555	12135	11816	10723	10974	10575	9487

Table 1.2.2.13**Mean Scale Scores by Ethnicity for Grade 11, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	389.45	379.63	345.7	303.38	346.63	362.78	382.83	357.65
Hispanic (of any Race)	N	75925	75296	78564	76449	70374	71850	69753	62641
Non-Hispanic American Indian	Mean	403.55	386.02	354.74	316.79	361.14	370.67	391.96	367.64
Non-Hispanic American Indian	N	1121	1195	1254	1173	1016	1146	1030	911
Non-Hispanic Asian	Mean	410.24	395.94	361.9	332.4	371.54	379.12	400.59	376.75
Non-Hispanic Asian	N	8546	8325	8637	8396	7802	7850	7780	6844
Non-Hispanic Black	Mean	400.26	386.66	346.96	322.59	361.69	367	391.1	365.34
Non-Hispanic Black	N	6714	6579	6899	6778	6192	6208	6063	5367
Non-Hispanic Multiracial	Mean	410.42	390.86	353.51	325.79	368.54	373.07	397.82	372.05
Non-Hispanic Multiracial	N	385	377	389	381	360	360	358	321
Non-Hispanic Pacific Islander	Mean	395.04	380.45	354.32	309.52	351.84	367.67	384.82	362.19
Non-Hispanic Pacific Islander	N	859	863	892	842	744	792	767	628
Non-Hispanic White	Mean	403.67	388.31	349.46	319.35	361.88	369.27	393.48	367.29
Non-Hispanic White	N	8769	8499	9028	8658	8002	8135	7904	7032
Unknown	Mean	384.86	376.28	335.49	295.8	340.33	355.85	379.17	350.85
Unknown	N	9256	9198	9657	9383	8501	8678	8376	7391

Table 1.2.2.14**Mean Scale Scores by Ethnicity for Grade 12, S602 Online**

Ethnicity	Statistic	Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
Hispanic (of any Race)	Mean	392.15	380.96	346.8	306.31	349.29	364.02	384.55	359.24
Hispanic (of any Race)	N	52950	52562	54703	53931	49903	50373	49024	44820
Non-Hispanic American Indian	Mean	405.9	388.06	354.98	320.45	364.16	372.03	393.58	369.35
Non-Hispanic American Indian	N	863	933	971	905	785	893	799	704
Non-Hispanic Asian	Mean	409.61	395	362.16	332.98	371.46	378.83	399.65	376.16
Non-Hispanic Asian	N	6624	6453	6689	6613	6176	6106	6050	5423
Non-Hispanic Black	Mean	399.91	386.72	346.48	324.06	362.11	366.75	391.02	365.38
Non-Hispanic Black	N	5623	5464	5782	5771	5312	5182	5058	4606
Non-Hispanic Multiracial	Mean	400.02	387.35	348.57	317.24	359.51	368.96	390.95	365.96
Non-Hispanic Multiracial	N	277	279	290	283	259	271	254	234
Non-Hispanic Pacific Islander	Mean	393.59	379.98	353.73	311.05	352	366.89	384.06	361.78
Non-Hispanic Pacific Islander	N	695	683	706	678	622	637	629	537
Non-Hispanic White	Mean	406.44	389.27	348.96	318.76	362.86	369.27	395.07	367.35
Non-Hispanic White	N	6828	6611	7064	6903	6371	6322	6142	5554
Unknown	Mean	389.85	378.73	339.19	302.89	346.56	359.1	382.46	355.21
Unknown	N	6990	6938	7243	7215	6604	6598	6416	5828

1.2.3 Correlations

Tables in this section show Pearson correlations among the four domain scale scores by grade-level cluster across all tiers, as well as the number of students included in each correlation. The results are presented by grade-level cluster. The pattern of domain correlations varied across clusters. In grade 1, Listening was correlated to Speaking and Writing; Reading was correlated to Writing. In cluster 2–3, Listening was mostly correlated to Speaking and Writing, and Reading was correlated to Listening. In clusters 4–5 and 6–8, Listening was correlated to Reading and Writing, and Reading was correlated to Listening and Writing. In cluster 9–12, the Listening and Reading domains were highly correlated, and the Listening, Reading, and Writing domains were correlated to the Speaking domain.

Table 1.2.3.1

Correlations Among Scale Scores: Grade 1, S602 Online

Domains	Pearson Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.412	0.605	0.58
Listening	N	213555	203660	196145	213505
Reading	Pearson Correlation	N/A	1	0.357	0.493
Reading	N	N/A	223101	204161	223044
Speaking	Pearson Correlation	N/A	N/A	1	0.529
Speaking	N	N/A	N/A	214405	214350
Writing	Pearson Correlation	N/A	N/A	N/A	1
Writing	N	N/A	N/A	N/A	235734

Table 1.2.3.2

Correlations Among Scale Scores: Grades 2–3, S602 Online

Domains	Pearson Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.591	0.651	0.642
Listening	N	436928	412025	405599	436798
Reading	Pearson Correlation	N/A	1	0.501	0.555
Reading	N	N/A	452156	418696	452015
Speaking	Pearson Correlation	N/A	N/A	1	0.628
Speaking	N	N/A	N/A	445591	445456
Writing	Pearson Correlation	N/A	N/A	N/A	1
Writing	N	N/A	N/A	N/A	484822

Table 1.2.3.3**Correlations Among Scale Scores: Grades 4–5, S602 Online**

Domains	Pearson Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.695	0.672	0.687
Listening	N	373317	344035	345642	348879
Reading	Pearson Correlation	N/A	1	0.569	0.682
Reading	N	N/A	374121	346684	351554
Speaking	Pearson Correlation	N/A	N/A	1	0.672
Speaking	N	N/A	N/A	377729	352488
Writing	Pearson Correlation	N/A	N/A	N/A	1
Writing	N	N/A	N/A	N/A	381687

Table 1.2.3.4**Correlations Among Scale Scores: Grades 6–8, S602 Online**

Domains	Pearson Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.68	0.655	0.67
Listening	N	435520	401324	398577	415319
Reading	Pearson Correlation	N/A	1	0.584	0.695
Reading	N	N/A	446485	408742	427020
Speaking	Pearson Correlation	N/A	N/A	1	0.68
Speaking	N	N/A	N/A	448077	426123
Writing	Pearson Correlation	N/A	N/A	N/A	1
Writing	N	N/A	N/A	N/A	469133

Table 1.2.3.5**Correlations Among Scale Scores: Grades 9–12, S602 Online**

Domains	Pearson Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.682	0.584	0.563
Listening	N	493338	451351	458146	467609
Reading	Pearson Correlation	N/A	1	0.562	0.554
Reading	N	N/A	489225	455598	465480
Speaking	Pearson Correlation	N/A	N/A	1	0.638
Speaking	N	N/A	N/A	501163	474213
Writing	Pearson Correlation	N/A	N/A	N/A	1
Writing	N	N/A	N/A	N/A	511940

1.3 Proficiency Level Results

The proficiency level results display the distribution of students' language proficiency levels by grade-level cluster and grade, within four domains (Listening, Reading, Writing, Speaking) and four composites (Oral, Literacy, Comprehension, Overall).

1.3.1 Domains

Tables in this section provide information on student performance by proficiency level (PL) for each test form, including the number and percentage of students whose performance placed them into each proficiency level, by domain.

The performance by domain was observed in the descending order of Listening, Reading, Speaking, and Writing. For Listening, a large percentage obtained Proficiency Level (PL) 6, especially in cluster 4–5 amounting to about 59%. The Reading domain had 3.6% to 11.6% in PL 6. For the Writing domain, fewer than 1% of students were in PL 5 and PL 6 together, except cluster 4–5 with 1.45% in PL 5 and 6 combined. In the Speaking domain, fewer than 2% were in PL 5 and PL 6, except cluster 4–5, which showed nearly 4.7% in both PL ranges combined.

1.3.1.1 Listening

1.3.1.1.1 By Cluster

Table 1.3.1.1.1

Proficiency Level by Cluster (Count): Listening, S602 Online

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	43121	17086	32258	13443	26003	81644	213555
2–3	75757	56974	106604	44894	58001	94698	436928
4–5	16831	26048	37471	17326	56165	219476	373317
6–8	32478	36691	90401	69480	79639	126831	435520
9–12	63483	61293	115402	103439	77326	72395	493338

Table 1.3.1.1.2

Proficiency Level by Cluster (Percent): Listening, S602 Online

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	20.19%	8.0%	15.11%	6.29%	12.18%	38.23%	100.0%
2–3	17.34%	13.04%	24.4%	10.27%	13.27%	21.67%	100.0%
4–5	4.51%	6.98%	10.04%	4.64%	15.04%	58.79%	100.0%
6–8	7.46%	8.42%	20.76%	15.95%	18.29%	29.12%	100.0%
9–12	12.87%	12.42%	23.39%	20.97%	15.67%	14.67%	100.0%

1.3.1.1.2 By Grade

Table 1.3.1.1.2.1

Proficiency Level by Grade (Count): Listening, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	43121	17086	32258	13443	26003	81644	213555
2	37757	30380	56073	23789	29093	44090	221182
3	38000	26594	50531	21105	28908	50608	215746
4	6508	12751	23510	8954	23719	128189	203631
5	10323	13297	13961	8372	32446	91287	169686
6	7174	10830	33291	22383	31493	35662	140833
7	10716	12503	29929	24142	29420	40983	147693
8	14588	13358	27181	22955	18726	50186	146994
9	17378	20474	38543	35122	24586	25334	161437
10	17061	16654	33987	27865	22774	21135	139476
11	16614	14275	24333	22238	19307	14808	111575
12	12430	9890	18539	18214	10659	11118	80850

Table 1.3.1.1.2.2

Proficiency Level by Grade (Percent): Listening, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	20.19%	8.0%	15.11%	6.29%	12.18%	38.23%	100.0%
2	17.07%	13.74%	25.35%	10.76%	13.15%	19.93%	100.0%
3	17.61%	12.33%	23.42%	9.78%	13.4%	23.46%	100.0%
4	3.2%	6.26%	11.55%	4.4%	11.65%	62.95%	100.0%
5	6.08%	7.84%	8.23%	4.93%	19.12%	53.8%	100.0%
6	5.09%	7.69%	23.64%	15.89%	22.36%	25.32%	100.0%
7	7.26%	8.47%	20.26%	16.35%	19.92%	27.75%	100.0%
8	9.92%	9.09%	18.49%	15.62%	12.74%	34.14%	100.0%
9	10.76%	12.68%	23.87%	21.76%	15.23%	15.69%	100.0%
10	12.23%	11.94%	24.37%	19.98%	16.33%	15.15%	100.0%
11	14.89%	12.79%	21.81%	19.93%	17.3%	13.27%	100.0%
12	15.37%	12.23%	22.93%	22.53%	13.18%	13.75%	100.0%

1.3.1.2 Reading

1.3.1.2.1 By Cluster

Table 1.3.1.2.1.1

Proficiency Level by Cluster (Count): Reading, S602 Online

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	46553	76336	42961	21927	20813	14511	223101
2-3	72873	103780	102897	63876	68147	40583	452156
4-5	71837	85716	68916	39059	70089	38504	374121
6-8	151175	125080	84947	29609	39590	16084	446485
9-12	96689	142847	102256	30291	60255	56887	489225

Table 1.3.1.2.1.2

Proficiency Level by Cluster (Percent): Reading, S602 Online

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	20.87%	34.22%	19.26%	9.83%	9.33%	6.5%	100.0%
2-3	16.12%	22.95%	22.76%	14.13%	15.07%	8.98%	100.0%
4-5	19.2%	22.91%	18.42%	10.44%	18.73%	10.29%	100.0%
6-8	33.86%	28.01%	19.03%	6.63%	8.87%	3.6%	100.0%
9-12	19.76%	29.2%	20.9%	6.19%	12.32%	11.63%	100.0%

1.3.1.2.2 By Grade

Table 1.3.1.2.2.1

Proficiency Level by Grade (Count): Reading, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	46553	76336	42961	21927	20813	14511	223101
2	28092	46760	64938	37064	38089	15794	230737
3	44781	57020	37959	26812	30058	24789	221419
4	33593	47961	36062	23252	40792	22610	204270
5	38244	37755	32854	15807	29297	15894	169851
6	48763	46159	26476	8248	11936	3911	145493
7	51304	43238	27970	10759	12972	5147	151390
8	51108	35683	30501	10602	14682	7026	149602
9	29821	47982	35512	10872	19791	16893	160871
10	24461	41647	28668	9019	17459	16845	138099
11	23722	30242	21356	7062	13736	14214	110332
12	18685	22976	16720	3338	9269	8935	79923

Table 1.3.1.2.2.2**Proficiency Level by Grade (Percent): Reading, S602 Online**

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	20.87%	34.22%	19.26%	9.83%	9.33%	6.5%	100.0%
2	12.17%	20.27%	28.14%	16.06%	16.51%	6.85%	100.0%
3	20.22%	25.75%	17.14%	12.11%	13.58%	11.2%	100.0%
4	16.45%	23.48%	17.65%	11.38%	19.97%	11.07%	100.0%
5	22.52%	22.23%	19.34%	9.31%	17.25%	9.36%	100.0%
6	33.52%	31.73%	18.2%	5.67%	8.2%	2.69%	100.0%
7	33.89%	28.56%	18.48%	7.11%	8.57%	3.4%	100.0%
8	34.16%	23.85%	20.39%	7.09%	9.81%	4.7%	100.0%
9	18.54%	29.83%	22.07%	6.76%	12.3%	10.5%	100.0%
10	17.71%	30.16%	20.76%	6.53%	12.64%	12.2%	100.0%
11	21.5%	27.41%	19.36%	6.4%	12.45%	12.88%	100.0%
12	23.38%	28.75%	20.92%	4.18%	11.6%	11.18%	100.0%

1.3.1.3 Writing*1.3.1.3.1 By Cluster***Table 1.3.1.3.1.1****Proficiency Level by Cluster (Count): Writing, S602 Online**

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	104126	80942	48257	2311	77	21	235734
2-3	86216	95959	257114	44606	884	43	484822
4-5	47986	36673	182565	108941	4877	645	381687
6-8	81213	90802	240477	56177	458	6	469133
9-12	82953	97017	253823	76771	1362	14	511940

Table 1.3.1.3.1.2**Proficiency Level by Cluster (Percent): Writing, S602 Online**

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	44.17%	34.34%	20.47%	0.98%	0.03%	0.01%	100.0%
2-3	17.78%	19.79%	53.03%	9.2%	0.18%	0.01%	100.0%
4-5	12.57%	9.61%	47.83%	28.54%	1.28%	0.17%	100.0%
6-8	17.31%	19.36%	51.26%	11.97%	0.1%	0.0%	100.0%
9-12	16.2%	18.95%	49.58%	15.0%	0.27%	0.0%	100.0%

1.3.1.3.2 By Grade

Table 1.3.1.3.2.1

Proficiency Level by Grade (Count): Writing, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	104126	80942	48257	2311	77	21	235734
2	49811	61995	123404	11937	102	8	247257
3	36405	33964	133710	32669	782	35	237565
4	27384	19750	106570	52170	1696	469	208039
5	20602	16923	75995	56771	3181	176	173648
6	27215	26995	85300	13279	112	0	152901
7	24459	39048	70196	25703	96	0	159502
8	29539	24759	84981	17195	250	6	156730
9	24039	36051	73808	34201	510	13	168622
10	18070	25574	84244	16052	609	1	144550
11	20777	24168	54771	15453	151	0	115320
12	20067	11224	41000	11065	92	0	83448

Table 1.3.1.3.2.2

Proficiency Level by Grade (Percent): Writing, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	44.17%	34.34%	20.47%	0.98%	0.03%	0.01%	100.0%
2	20.15%	25.07%	49.91%	4.83%	0.04%	0.0%	100.0%
3	15.32%	14.3%	56.28%	13.75%	0.33%	0.01%	100.0%
4	13.16%	9.49%	51.23%	25.08%	0.82%	0.23%	100.0%
5	11.86%	9.75%	43.76%	32.69%	1.83%	0.1%	100.0%
6	17.8%	17.66%	55.79%	8.68%	0.07%	0.0%	100.0%
7	15.33%	24.48%	44.01%	16.11%	0.06%	0.0%	100.0%
8	18.85%	15.8%	54.22%	10.97%	0.16%	0.0%	100.0%
9	14.26%	21.38%	43.77%	20.28%	0.3%	0.01%	100.0%
10	12.5%	17.69%	58.28%	11.1%	0.42%	0.0%	100.0%
11	18.02%	20.96%	47.49%	13.4%	0.13%	0.0%	100.0%
12	24.05%	13.45%	49.13%	13.26%	0.11%	0.0%	100.0%

1.3.1.4 Speaking

1.3.1.4.1 By Cluster

Table 1.3.1.4.1.1

Proficiency Level by Cluster (Count): Speaking, S602 Online

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	65687	65432	68063	13848	1264	111	214405
2-3	102259	119598	164552	53371	5269	542	445591
4-5	66824	82440	117429	93370	15447	2219	377729
6-8	138443	81550	160140	65473	2157	314	448077
9-12	198673	109040	178141	14460	639	210	501163

Table 1.3.1.4.1.2

Proficiency Level by Cluster (Percent): Speaking, S602 Online

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	30.64%	30.52%	31.75%	6.46%	0.59%	0.05%	100.0%
2-3	22.95%	26.84%	36.93%	11.98%	1.18%	0.12%	100.0%
4-5	17.69%	21.83%	31.09%	24.72%	4.09%	0.59%	100.0%
6-8	30.9%	18.2%	35.74%	14.61%	0.48%	0.07%	100.0%
9-12	39.64%	21.76%	35.55%	2.89%	0.13%	0.04%	100.0%

1.3.1.4.2 By Grade

Table 1.3.1.4.2.1

Proficiency Level by Grade (Count): Speaking, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	65687	65432	68063	13848	1264	111	214405
2	55930	72667	69316	25019	2526	149	225607
3	46329	46931	95236	28352	2743	393	219984
4	31099	48286	66012	51400	8182	1166	206145
5	35725	34154	51417	41970	7265	1053	171584
6	40496	29992	56023	18680	541	35	145767
7	49536	28347	47069	25166	1111	93	151322
8	48411	23211	57048	21627	505	186	150988
9	73302	32817	53618	5547	133	45	165462
10	55206	28272	53781	3866	165	52	141342
11	40991	22594	44431	3758	220	66	112060
12	29174	25357	26311	1289	121	47	82299

Table 1.3.1.4.2.2**Proficiency Level by Grade (Percent): Speaking, S602 Online**

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	30.64%	30.52%	31.75%	6.46%	0.59%	0.05%	100.0%
2	24.79%	32.21%	30.72%	11.09%	1.12%	0.07%	100.0%
3	21.06%	21.33%	43.29%	12.89%	1.25%	0.18%	100.0%
4	15.09%	23.42%	32.02%	24.93%	3.97%	0.57%	100.0%
5	20.82%	19.91%	29.97%	24.46%	4.23%	0.61%	100.0%
6	27.78%	20.58%	38.43%	12.81%	0.37%	0.02%	100.0%
7	32.74%	18.73%	31.11%	16.63%	0.73%	0.06%	100.0%
8	32.06%	15.37%	37.78%	14.32%	0.33%	0.12%	100.0%
9	44.3%	19.83%	32.41%	3.35%	0.08%	0.03%	100.0%
10	39.06%	20.0%	38.05%	2.74%	0.12%	0.04%	100.0%
11	36.58%	20.16%	39.65%	3.35%	0.2%	0.06%	100.0%
12	35.45%	30.81%	31.97%	1.57%	0.15%	0.06%	100.0%

1.3.2 Composites

This section presents students' performance in the four composite areas, by proficiency level. Tables show the proficiency levels by student counts and percentages for each grade and grade-level cluster.

The observed order of performance of composite domains by percentages in PL 5 and 6, in descending order, was Comprehension, Oral, Overall, and Literacy.

1.3.2.1 Oral Composite*1.3.2.1.1 By Cluster***Table 1.3.2.1.1.1****Proficiency Level by Cluster (Count): Oral, S602 Online**

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	44045	35361	59286	39637	15951	1865	196145
2-3	72759	87571	131493	84806	26107	2863	405599
4-5	32982	34247	64878	107979	74100	31456	345642
6-8	64200	66056	118123	116450	29848	3900	398577
9-12	111040	92147	166396	78935	8549	1079	458146

Table 1.3.2.1.1.2**Proficiency Level by Cluster (Percent): Oral, S602 Online**

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	22.46%	18.03%	30.23%	20.21%	8.13%	0.95%	100.0%
2–3	17.94%	21.59%	32.42%	20.91%	6.44%	0.71%	100.0%
4–5	9.54%	9.91%	18.77%	31.24%	21.44%	9.1%	100.0%
6–8	16.11%	16.57%	29.64%	29.22%	7.49%	0.98%	100.0%
9–12	24.24%	20.11%	36.32%	17.23%	1.87%	0.24%	100.0%

*1.3.2.1.2 By Grade***Table 1.3.2.1.2.1****Proficiency Level by Grade (Count): Oral, S602 Online**

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	44045	35361	59286	39637	15951	1865	196145
2	36328	49528	66896	38371	11777	1184	204084
3	36431	38043	64597	46435	14330	1679	201515
4	14523	19408	34756	57100	42439	19839	188065
5	18459	14839	30122	50879	31661	11617	157577
6	16657	22922	41201	37845	9020	1115	128760
7	21843	22354	39454	39531	10092	1279	134553
8	25700	20780	37468	39074	10736	1506	135264
9	36087	30700	52673	27325	2659	349	149793
10	30450	26047	46551	23411	2563	308	129330
11	25695	20488	37225	17160	2154	269	102991
12	18808	14912	29947	11039	1173	153	76032

Table 1.3.2.1.2.2**Proficiency Level by Grade (Percent): Oral, S602 Online**

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	22.46%	18.03%	30.23%	20.21%	8.13%	0.95%	100.0%
2	17.8%	24.27%	32.78%	18.8%	5.77%	0.58%	100.0%
3	18.08%	18.88%	32.06%	23.04%	7.11%	0.83%	100.0%
4	7.72%	10.32%	18.48%	30.36%	22.57%	10.55%	100.0%
5	11.71%	9.42%	19.12%	32.29%	20.09%	7.37%	100.0%
6	12.94%	17.8%	32.0%	29.39%	7.01%	0.87%	100.0%
7	16.23%	16.61%	29.32%	29.38%	7.5%	0.95%	100.0%
8	19.0%	15.36%	27.7%	28.89%	7.94%	1.11%	100.0%
9	24.09%	20.49%	35.16%	18.24%	1.78%	0.23%	100.0%
10	23.54%	20.14%	35.99%	18.1%	1.98%	0.24%	100.0%
11	24.95%	19.89%	36.14%	16.66%	2.09%	0.26%	100.0%
12	24.74%	19.61%	39.39%	14.52%	1.54%	0.2%	100.0%

1.3.2.2 Literacy Composite*1.3.2.2.1 By Cluster***Table 1.3.2.2.1.1****Proficiency Level by Cluster (Count): Literacy, S602 Online**

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	84731	79732	47278	9044	1893	366	223044
2-3	74390	104471	192944	71962	7506	742	452015
4-5	56691	48454	126458	96912	20055	2984	351554
6-8	96505	107561	171637	48365	2815	137	427020
9-12	69869	119428	189461	74370	11841	511	465480

Table 1.3.2.2.1.2**Proficiency Level by Cluster (Percent): Literacy, S602 Online**

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	37.99%	35.75%	21.2%	4.05%	0.85%	0.16%	100.0%
2-3	16.46%	23.11%	42.69%	15.92%	1.66%	0.16%	100.0%
4-5	16.13%	13.78%	35.97%	27.57%	5.7%	0.85%	100.0%
6-8	22.6%	25.19%	40.19%	11.33%	0.66%	0.03%	100.0%
9-12	15.01%	25.66%	40.7%	15.98%	2.54%	0.11%	100.0%

1.3.2.2.2 By Grade

Table 1.3.2.2.1

Proficiency Level by Grade (Count): Literacy, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	84731	79732	47278	9044	1893	366	223044
2	37621	60850	99894	29800	2222	262	230649
3	36769	43621	93050	42162	5284	480	221366
4	29675	26263	70745	52042	10709	1811	191245
5	27016	22191	55713	44870	9346	1173	160309
6	31360	37552	58437	11215	505	36	139105
7	31205	37883	58646	16110	981	55	144880
8	33940	32126	54554	21040	1329	46	143035
9	22505	35986	63197	26793	4060	271	152812
10	17233	33310	54407	22556	3595	166	131267
11	16267	27675	42070	16017	2927	63	105019
12	13864	22457	29787	9004	1259	11	76382

Table 1.3.2.2.2

Proficiency Level by Grade (Percent): Literacy, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	37.99%	35.75%	21.2%	4.05%	0.85%	0.16%	100.0%
2	16.31%	26.38%	43.31%	12.92%	0.96%	0.11%	100.0%
3	16.61%	19.71%	42.03%	19.05%	2.39%	0.22%	100.0%
4	15.52%	13.73%	36.99%	27.21%	5.6%	0.95%	100.0%
5	16.85%	13.84%	34.75%	27.99%	5.83%	0.73%	100.0%
6	22.54%	27.0%	42.01%	8.06%	0.36%	0.03%	100.0%
7	21.54%	26.15%	40.48%	11.12%	0.68%	0.04%	100.0%
8	23.73%	22.46%	38.14%	14.71%	0.93%	0.03%	100.0%
9	14.73%	23.55%	41.36%	17.53%	2.66%	0.18%	100.0%
10	13.13%	25.38%	41.45%	17.18%	2.74%	0.13%	100.0%
11	15.49%	26.35%	40.06%	15.25%	2.79%	0.06%	100.0%
12	18.15%	29.4%	39.0%	11.79%	1.65%	0.01%	100.0%

1.3.2.3 Comprehension Composite

1.3.2.3.1 By Cluster

Table 1.3.2.3.1.1

Proficiency Level by Cluster (Count): Comprehension, S602 Online

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	30073	48399	53740	22295	28737	20416	203660
2-3	52837	97768	102069	52075	62038	45238	412025
4-5	29814	54562	56431	43653	74510	85065	344035
6-8	64348	107043	90148	53889	56827	29069	401324
9-12	61505	120173	104759	56994	60184	47736	451351

Table 1.3.2.3.1.2

Proficiency Level by Cluster (Percent): Comprehension, S602 Online

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	14.77%	23.76%	26.39%	10.95%	14.11%	10.02%	100.0%
2-3	12.82%	23.73%	24.77%	12.64%	15.06%	10.98%	100.0%
4-5	8.67%	15.86%	16.4%	12.69%	21.66%	24.73%	100.0%
6-8	16.03%	26.67%	22.46%	13.43%	14.16%	7.24%	100.0%
9-12	13.63%	26.63%	23.21%	12.63%	13.33%	10.58%	100.0%

1.3.2.3.2 By Grade

Table 1.3.2.3.2.1

Proficiency Level by Grade (Count): Comprehension, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	30073	48399	53740	22295	28737	20416	203660
2	17851	52591	57988	30210	30495	19505	208640
3	34986	45177	44081	21865	31543	25733	203385
4	11149	30728	31005	22893	40948	50566	187289
5	18665	23834	25426	20760	33562	34499	156746
6	16560	39593	33125	16736	16875	6893	129782
7	22012	36008	29870	18731	18832	10414	135867
8	25776	31442	27153	18422	21120	11762	135675
9	16462	40734	35239	19919	20735	14312	147401
10	15692	33267	30633	16674	17117	14164	127547
11	16347	26230	22417	11590	13506	11941	102031
12	13004	19942	16470	8811	8826	7319	74372

Table 1.3.2.3.2.2**Proficiency Level by Grade (Percent): Comprehension, S602 Online**

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	14.77%	23.76%	26.39%	10.95%	14.11%	10.02%	100.0%
2	8.56%	25.21%	27.79%	14.48%	14.62%	9.35%	100.0%
3	17.2%	22.21%	21.67%	10.75%	15.51%	12.65%	100.0%
4	5.95%	16.41%	16.55%	12.22%	21.86%	27.0%	100.0%
5	11.91%	15.21%	16.22%	13.24%	21.41%	22.01%	100.0%
6	12.76%	30.51%	25.52%	12.9%	13.0%	5.31%	100.0%
7	16.2%	26.5%	21.98%	13.79%	13.86%	7.66%	100.0%
8	19.0%	23.17%	20.01%	13.58%	15.57%	8.67%	100.0%
9	11.17%	27.63%	23.91%	13.51%	14.07%	9.71%	100.0%
10	12.3%	26.08%	24.02%	13.07%	13.42%	11.1%	100.0%
11	16.02%	25.71%	21.97%	11.36%	13.24%	11.7%	100.0%
12	17.49%	26.81%	22.15%	11.85%	11.87%	9.84%	100.0%

1.3.2.4 Overall Composite*1.3.2.4.1 By Cluster***Table 1.3.2.4.1.1****Proficiency Level by Cluster (Count): Overall, S602 Online**

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	50185	65500	58706	10889	2274	230	187784
2-3	62957	88285	157858	67051	7693	323	384167
4-5	38444	37067	92494	102867	29800	3738	304410
6-8	66717	79118	145816	62652	3871	172	358346
9-12	73922	94792	167734	61708	6748	252	405156

Table 1.3.2.4.1.2**Proficiency Level by Cluster (Percent): Overall, S602 Online**

Cluster	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	26.72%	34.88%	31.26%	5.8%	1.21%	0.12%	100.0%
2-3	16.39%	22.98%	41.09%	17.45%	2.0%	0.08%	100.0%
4-5	12.63%	12.18%	30.38%	33.79%	9.79%	1.23%	100.0%
6-8	18.62%	22.08%	40.69%	17.48%	1.08%	0.05%	100.0%
9-12	18.25%	23.4%	41.4%	15.23%	1.67%	0.06%	100.0%

1.3.2.4.2 By Grade

Table 1.3.2.4.2.1

Proficiency Level by Grade (Count): Overall, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	50185	65500	58706	10889	2274	230	187784
2	31170	51162	81153	27005	2708	133	193331
3	31787	37123	76705	40046	4985	190	190836
4	18836	20117	50849	55277	17164	2429	164672
5	19608	16950	41645	47590	12636	1309	139738
6	19087	27759	51495	16475	758	49	115623
7	22651	26629	48701	21615	1341	69	121006
8	24979	24730	45620	24562	1772	54	121717
9	23681	29044	55599	21224	2397	135	132080
10	19502	25903	47990	18783	1985	72	114235
11	17199	21523	37031	13777	1567	38	91135
12	13540	18322	27114	7924	799	7	67706

Table 1.3.2.4.2.2

Proficiency Level by Grade (Percent): Overall, S602 Online

Grade	PL 1	PL 2	PL 3	PL 4	PL 5	PL 6	Total
1	26.72%	34.88%	31.26%	5.8%	1.21%	0.12%	100.0%
2	16.12%	26.46%	41.98%	13.97%	1.4%	0.07%	100.0%
3	16.66%	19.45%	40.19%	20.98%	2.61%	0.1%	100.0%
4	11.44%	12.22%	30.88%	33.57%	10.42%	1.48%	100.0%
5	14.03%	12.13%	29.8%	34.06%	9.04%	0.94%	100.0%
6	16.51%	24.01%	44.54%	14.25%	0.66%	0.04%	100.0%
7	18.72%	22.01%	40.25%	17.86%	1.11%	0.06%	100.0%
8	20.52%	20.32%	37.48%	20.18%	1.46%	0.04%	100.0%
9	17.93%	21.99%	42.09%	16.07%	1.81%	0.1%	100.0%
10	17.07%	22.68%	42.01%	16.44%	1.74%	0.06%	100.0%
11	18.87%	23.62%	40.63%	15.12%	1.72%	0.04%	100.0%
12	20.0%	27.06%	40.05%	11.7%	1.18%	0.01%	100.0%

2. Analysis of Domains

The measurement model that forms the basis of the analysis for the development of ACCESS for ELLs is the Rasch measurement model (Wright & Stone, 1979). Additional information on its use in the development of the ACCESS for ELLs assessment program is available in WIDA Technical Report No. 1, *Development and Field Test of ACCESS for ELLs* (Kenyon, 2006). The original ACCESS test developers used Rasch measurement principles, and in that sense, the Rasch model guided all decisions throughout the development of the assessment and was not just a tool for the statistical analysis of the data. Thus, for example, data based on Rasch fit statistics guided the inclusion, revision, or deletion of items during the development and field testing of the test forms and will continue to guide the refinement and further development of the test. All Rasch analyses are conducted using the Rasch measurement software program *Winsteps* (Linacre, 2006).

For Listening and Reading, the dichotomous Rasch model was used as the measurement model. Mathematically, the measurement model may be presented as

$$\log\left(\frac{P_{ni1}}{P_{ni0}}\right) = B_n - D_i$$

where

P_{ni1} = probability of providing a correct response "1" by student "n" to item "i"

P_{ni0} = probability of providing an incorrect response "0" by student "n" to item "i"

B_n = ability of student "n"

D_i = difficulty of item "i"

When the probability of a student providing a correct answer to an item equals the probability of a student providing an incorrect answer (i.e., 50% probability of getting it right and 50% probability of getting it wrong), P_{ni1}/P_{ni0} is equal to 1. The log of 1 is 0. This is the point at which a student's ability equals the difficulty of an item. For example, a student whose ability estimate is 1.56 on the Rasch logit scale encountering an item whose difficulty is 1.56 on the Rasch logit scale would have a 50% probability of providing a correct answer to that item.

The Rasch model was also used to score polytomous tasks. The Writing and Speaking tasks used a Rasch-grouped rating scale model, which is an extension of Andrich's rating scale model (Andrich, 1978). Mathematically, this can be represented as

$$\log\left(\frac{P_{ngik}}{P_{ngi(k-1)}}\right) = \beta_n - D_{gi} - F_{gk}$$

where

P_{ngik} = probability of student "n" on task "i" receiving a rating at level "k" on rating scale "g"

$P_{ngi(k-1)}$ = probability of student "n" on task "i" receiving a rating at level "k - 1" on rating scale "g" (i.e., the next lowest rating)

β_n = ability of student "n"

D_{gi} = difficulty of task "i" specific to rating scale "g"

F_{gk} = step calibration value of category "k" relative to category 'k - 1' on rating scale "g"

The subscript "g" is a group index specifying the group of tasks to which task "i" belongs. It also identifies the rating scale that was used for the group of tasks. There is only one rating scale ($g = 1$) in the Writing domain and two grouped rating scales ($g = 2$) in the Speaking domain. As with the dichotomous Rasch model, there is an item difficulty parameter (D_{gi}) for each item for rating scale "g" modeled by the Rasch rating scale model (Andrich, 1978). In addition, there is a step calibration value or *step measure* (F_{gk}) that corresponds to the location on the latent variable where the probability of being observed in the "k" and "k - 1" category for rating scale "g" is equal, relative to the difficulty measure of the task. The step measures are also the points where adjacent category probability "k - 1" and "k" curves for rating scale "g" intercept. All tasks that belong to the same rating scale group have the same step measures. As described in Part 1, Section 4.3, ratings on the ACCESS Writing Scoring Scale range from 0, 1, 1+, ..., 6, and the possible raw scores range from 0 to 9. Writing raters use this scoring scale for all Writing tasks. We model all other Writing tasks using a single rating scale with possible raw scores of 0 to 9.

In 2015–2016, with the transition to Online ACCESS, CAL conducted a Writing scaling study. Detailed information about the derivation of the Writing rating scale as well as the psychometric properties of the Writing rating scale are available in the 2016 scaling report (Center for Applied Linguistics, 2017). In 2019–2020, we redesigned the Writing test to allow for embedded field testing, reducing the number of operational tasks from three to two. For details on how we retained the 2016 rating scale parameters and maintained the Writing score scale, see Center for Applied Linguistics (2019).

For Speaking, we model PL 1 tasks as a group on a 0–2 scale, and PL 3 and PL 5 tasks as a group on a 0–4 scale (see Part 1, Section 4.4). We conducted a study in the summer of 2016 to reconstruct the logit scales, and detailed information about the derivation as well as the psychometric properties of Speaking rating scales are available in the scaling report (Center for Applied Linguistics, 2017).

Scale scores are calculated by transforming the student ability estimate via a scaling equation. The following scaling equations convert ability measures in logits to scale scores:

- L: (Ability Measure in Logits * 37.571) + 316.637
- R: (Ability Measure in Logits * 26.000) + 323.272
- W: (Ability Measure in Logits * 26.851) + 303.332
- S: (Ability Measure in Logits * 29.248) + 265.076

In the domains of Listening and Reading, we established the current ACCESS scale for the original paper-only version of the test and maintained this scale through the transition to an online- and paper-delivered test in the 2015–2016 school year (Series 400). Evidence for scale maintenance in the transitional year is described elsewhere (Center for Applied Linguistics, 2016). In the domains of Writing and Speaking, we conducted a study in the summer of 2016 to reconstruct the logit scale (Center for Applied Linguistics, 2017).

PL scores are interpretations of these scale scores in terms of the proficiency levels described in the WIDA ELD Standards. These interpretations derive from a series of standard-setting studies, in which educators reviewed evidence from the test, either in the form of items for the selected response sections (Listening and Reading) or student portfolios for the constructed response sections (Writing and Speaking), to establish cut scores between the proficiency levels. The first standard-setting study for ACCESS took place in 2005; it established cut scores for all four domains by grade-level cluster (Kenyon, 2006). The second cut score study took place in 2007; it established cut scores for all four domains by grade level (Kenyon, Ryu, & MacGregor, 2013). These cut scores were used to derive proficiency level scores through the 2015–2016 administration (Series 400) of ACCESS for ELLs. WIDA and CAL conducted a third cut score study in the summer of 2016 (Cook & MacGregor, 2017). The purpose of this study was to re-examine cut scores for each of the proficiency levels in light of the migration from the paper-and-pencil-only assessment to both online and paper delivery, the revision of the Speaking test, and the influence of college- and career-ready standards. These new cut scores were first used for ACCESS Series 401 (2016–2017 school year).

A proficiency level score consists of a two-digit decimal number (e.g., 4.5). The first digit represents the student’s overall proficiency level range based on the student’s scale score. The number to the right of the decimal point is an indication of the proportion of the range between cut scores that the student’s scale score represents. A score of 4.5, for example, tells us that the student is in PL 4 and that the student’s scale score is halfway between the cut scores for PL 4 and PL 5.

Unlike the scale scores, which form an interval scale and are continuous across grades from kindergarten to grade 12, PL scores are dependent upon the grade a student was in when the student took the assessment. For example, a score of 350 in Listening would be interpreted as a PL score of 5.8 for a grade 2 student, a 3.8 for a grade 5 student, a 3.1 for a grade 8 student, and a 2.3 for a grade 12 student.

Because the bands between cut scores on the score scale vary in width, PL scores do not form an interval scale. Only scale scores should be used as interval measures. PL scores are at even intervals within a grade and proficiency level (e.g., in grade 3, the distance between 3.1 and 3.2 is the same as the distance between 3.7 and 3.8), but they do not form an interval scale across proficiency levels.

2.1 Complete Item or Task Analysis and Summary

The tables in this section provide information on the psychometric qualities of the items and tasks. We provide values for item or task difficulties in logits, the number of items or tasks on the form, the average *p-value* (for forms with selected response items), and the Rasch model fit statistics. For Writing and Speaking, we also provide raw score distributions by task.

Tables in this section have either two parts (in the case of Listening and Reading) or three parts (in the case of Writing and Speaking). The first part of the table gives a summary of the total set of items or tasks on the form. The second part provides statistics pertaining to the individual

items or tasks, and the third part (for Writing and Speaking only) expresses raw score distributions by task.

For Listening and Reading, items form a pool for the multistage adaptive tests, and tables in this section provide information on every item in the grade-level cluster. For Writing, separate tables are provided for Tier A and Tier B/C forms, by grade-level cluster. For Speaking, which has tasks that are shared between Tier A and Tier B/C, there is one table for each grade-level cluster, which provides information on every task in the grade-level cluster.

All Rasch analyses were conducted using the Rasch measurement software program *Winsteps* 5.2.4.0 (Linacre, 2006). When speaking of the measure of student ability, we use the term “*ability measure*” (rather than “*theta*”, used commonly when discussing models based on item response theory). When speaking of the measure of how hard an item is, we use the term “*item difficulty measure*” (rather than “*b parameter*”, used commonly when discussing models based on item response theory). “*Step measures*” refer to the calibration of the steps in the Rasch rating scale model previously presented. All three measures (ability, difficulty, and step) are expressed in terms of Rasch logits, which then are converted into scores on the ACCESS score scale for reporting purposes.

Fit statistics for the Rasch model are calculated by comparing the observed empirical data with the data that the Rasch model would be expected to produce if the data fit the model perfectly. Outfit mean square statistics for items and tasks are influenced by outlier responses for machine-scored dichotomous items or outlier ratings for rater-scored performance tasks. For example, a difficult item that some low-ability students get correct—for reasons unknown—will have a high outfit mean square statistic. Similarly, an easy item that some high-ability students get wrong will also have a high outfit mean square statistic. Infit mean square statistics are influenced by unexpected patterns of students’ responses and ratings on items and tasks that are roughly targeted for them and generally indicate a more serious measurement problem. The expectation for both statistics is 1.00, and values near 1.00 are not of great concern. Values less than 1.00 indicate that the response and rating patterns are too predictable and thus redundant, or the model is overfitting the data, but are not of great concern. High values are of greater concern.

Linacre (2002b) provided more guidance on how to interpret these statistics for dichotomous items. He wrote:

- Values greater than 2.0 “distort or degrade the measurement system.” [Note: We interpret “degrade” here in the sense of lowering the quality of the measurement system.]
- Values between 1.5 and 2.0 are “unproductive for construction of measurement, but not degrading.”
- Values between 0.5 and 1.5 should be considered “productive for measurement.”
- Values below 0.5 are “less productive for measurement, but not degrading.”

Linacre also stated in his guidance that infit problems are more serious to the construction of measurement than outfit problems.

Because we follow conservative guidelines in the development of ACCESS for ELLs, it is desired that the dichotomous items on the test forms have mean square fit statistics in the range of 0.5 to 1.5; and thus, they fit the range that is “productive for measurement” according to the aforementioned guidelines. The percentages of dichotomous items which have mean square statistics within this range are included in the following subsections, by domain.

Since performance tasks are constructed and scored very differently from dichotomous items, it is not as straightforward to apply this same guidance to interpret these fit statistics for performance tasks that raters scored polytomously on a rubric scale. We design some performance tasks to elicit a restricted range of performances (for example, very easy tasks where we expect that most students will get the highest rating), and these tasks can cause the model to predict the data too well (overfitting). Conversely, when raters score performance tasks using a very wide rubric scale such as the ACCESS for ELLs Writing rubric, sometimes unmodeled noise or other sources of variance in the ratings of the students’ responses to the task will cause the model to underpredict those ratings (underfitting). Overall, for ACCESS for ELLs performance tasks, overfitting is more common than underfitting. Underfitting indicates that the task is less productive for measurement, but, according to Linacre (2002b), including the rating of the student’s performance on the task when calculating that student’s score does not degrade the measurement of the student’s performance.

The first section of the Complete Item/Task Analysis and Summary table provides information about the total set of items or tasks and includes the item type (selected response or constructed response), the average item difficulty measure (in logits), the number of items, the average *p-value* (for Listening and Reading only), the average infit mean square statistic, and the average outfit mean square statistic.

The second section of these tables presents results from the analyses of all the items or tasks on the test form. The first column in this section provides the unique item name. The second column presents the item or task difficulty measure, in logits. The third column indicates whether the item or task served as an anchor item or task, used to link score scales between series (see Section 2.7 for details), or is a dichotomously scored item (Listening and Reading). The fourth column shows the *p-value* (percentage of correct answers on that item). The final two columns show the Rasch fit statistics for the item or task. Folders with items that have fit statistics greater than 2.0 are evaluated by the test development team to determine whether and when the folders can be refreshed in the next test refreshment cycle.

In addition, the Writing and Speaking tables have a section at the bottom of the table that provides raw score distributions by task.

The results show that all items and tasks have infit mean square statistics less than 2.0 (which is the item selection and evaluation criteria) for all grade-level clusters and domains, indicating that the items and tasks provide trustworthy measures of ability for those students whose ability measures are in the region of the ability distribution that the items and tasks are targeting. As discussed earlier, the outfit mean square statistic is sensitive to outlier responses and scores that are not in the region of the ability distribution that the items and tasks are targeting. There are two items in the grades 2–3 Listening test that show outfit mean square statistics greater than 2.0. For the most part, these are very easy items, suggesting that there

might be some high-ability students getting these items incorrect and causing the outfit mean square statistics to be inflated.

All items in the Listening and Reading domains have infit mean square statistics between 0.5 and 1.5. All items in the Listening clusters 4–5 and 9–12, and all Reading clusters except cluster 1 have outfit mean square statistics that fall between 0.5 and 1.5. Listening clusters 1, 2–3, 6–8 and Reading cluster 1 have slightly lower outfit mean square statistics, with 98%, 94%, 98%, and 99% falling between 0.5 and 1.5, respectively.

2.1.1 Listening

2.1.1.1 Grade 1

Table 2.1.1.1

Complete Item Analysis and Summary: List 1 S602 Online

Table 2.1.1.1

Complete Item Analysis and Summary: List 1 S602 Online

[illegible]

Name	Item Difficulty (in logits)	Anchored?	P-value	Fit Statistics	
				Infit Mns q	Outfit Mns q
Information withheld due to confidentiality requirements.					

2.1.1.2 Grades 2–3

Table 2.1.1.2

Complete Item Analysis and Summary: List 2–3 S602 Online

Table 2.1.1.2

Complete Item Analysis and Summary: List 2-3 S602 Online

Item Type	Average Item Difficulty (in logits)	No. of Items	Average P-value	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.					

2.1.1.3 Grades 4–5

Table 2.1.1.3

Complete Item Analysis and Summary: List 4–5 S602 Online

Table 2.1.1.3

Complete Item Analysis and Summary: List 4-5 S602 Online

Item Type	Average Item Difficulty (in logits)	No. of Items	Average P-value	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.					

2.1.1.4 Grades 6–8

Table 2.1.1.4

Complete Item Analysis and Summary: List 6–8 S602 Online

Table 2.1.1.4

Complete Item Analysis and Summary: List 6-8 S602 Online

Item Type	Average Item Difficulty (in logits)	No. of Items	Average P-value	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.					

Name	Item Difficulty (in logits)	Anchored?	P-value	Fit Statistics	
				Infit Mnsq	Outfit Mnsq
Information withheld due to confidentiality requirements.					

2.1.1.5 Grades 9–12

Table 2.1.1.5

Complete Item Analysis and Summary: List 9–12 S602 Online

Table 2.1.1.5

Complete Item Analysis and Summary: List 9-12 S602 Online

Item Type	Average Item Difficulty (in logits)	No. of Items	Average P-value	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.					

Name	Item Difficulty (in logits)	Anchored?	P-value	Fit Statistics	
				Infit Mnsq	Outfit Mnsq
Information withheld due to confidentiality requirements.					

2.1.2 Reading

2.1.2.1 Grade 1

Table 2.1.2.1

Complete Item Analysis and Summary: Read 1 S602 Online

Table 2.1.2.1

Complete Item Analysis and Summary: Read 1 S602 Online

Item Type	Average Item Difficulty (in logits)	No. of Items	Average P-value	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.					

Name	Item Difficulty (in logits)	Anchored?	P-value	Fit Statistics	
				Infit Mnsq	Outfit Mnsq
Information withheld due to confidentiality requirements.					

2.1.2.2 Grades 2–3

Table 2.1.2.2

Complete Item Analysis and Summary: Read 2–3 S602 Online

Table 2.1.2.2

Complete Item Analysis and Summary: Read 2-3 S602 Online

Item Type	Average Item Difficulty (in logits)	No. of Items	Average P-value	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.					

2.1.2.3 Grades 4–5

Table 2.1.2.3

Complete Item Analysis and Summary: Read 4–5 S602 Online

Table 2.1.2.3

Complete Item Analysis and Summary: Read 4-5 S602 Online

Item Type	Average Item Difficulty (in logits)	No. of Items	Average P-value	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.					

Name	Item Difficulty (in logits)	Anchored?	P-value	Fit Statistics	
				Infit Mnsq	Outfit Mnsq
Information withheld due to confidentiality requirements.					

2.1.2.4 Grades 6–8

Table 2.1.2.4

Complete Item Analysis and Summary: Read 6–8 S602 Online

Table 2.1.2.4

Complete Item Analysis and Summary: Read 6-8 S602 Online

Item Type	Average Item Difficulty (in logits)	No. of Items	Average P-value	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.					

Name	Item Difficulty (in logits)	Anchored?	P-value	Fit Statistics	
				Infit Mnsq	Outfit Mnsq
Information withheld due to confidentiality requirements.					
2					
2					
2					
2					
3					
3					
3					
3					
3					
3					
3					
3					
3					
3					
3					
4					
4					
4					
4					
4					
4					
4					
4					
4					
5					
5					
5					
5					
5					
5					
5					
5					
5					
5					
5					
6					

2.1.2.5 Grades 9–12

Table 2.1.2.5

Complete Item Analysis and Summary: Read 9–12 S602 Online

Table 2.1.2.5
Complete Item Analysis and Summary: Read 9–12 S602 Online

Item Type	Average Item Difficulty (in logits)	No. of Items	Average P-value	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.					
1					
2					
3					
4					
5					
6					
7					
8					
9					
1					
1					
1					
1					
1					
1					
1					
1					
1					
2					
2					
2					
2					
2					
2					

2.1.3 Writing

2.1.3.1 Grade 1

Table 2.1.3.1.1

Complete Task Analysis and Summary: Writ 1 A S602 Online

Table 2.1.3.1.1

Complete Task Analysis and Summary: Writ 1 A S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

Table 2.1.3.1.2

Complete Task Analysis and Summary: Writ 1 B/C S602 Online

Table 2.1.3.1.2

Complete Task Analysis and Summary: Writ 1 B/C S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

2.1.3.2 Grades 2–3

Table 2.1.3.2.1

Complete Task Analysis and Summary: Writ 2–3 A S602 Online

Table 2.1.3.2.1

Complete Task Analysis and Summary: Writ 2-3 A S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

Table 2.1.3.2.2

Complete Task Analysis and Summary: Writ 2–3 B/C S602 Online

Table 2.1.3.2.2

Complete Task Analysis and Summary: Writ 2-3 B/C S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

2.1.3.3 Grades 4–5

Table 2.1.3.3.1

Complete Task Analysis and Summary: Writ 4–5 A S602 Online

Table 2.1.3.3.1

Complete Task Analysis and Summary: Writ 4-5 A S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

Table 2.1.3.3.2

Complete Task Analysis and Summary: Writ 4-5 B/C S602 Online

Table 2.1.3.3.2

Complete Task Analysis and Summary: Writ 4-5 B/C S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

2.1.3.4 Grades 6–8

Table 2.1.3.4.1

Complete Task Analysis and Summary: Writ 6–8 A S602 Online

Table 2.1.3.4.1

Complete Task Analysis and Summary: Writ 6-8 A S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

Table 2.1.3.4.2

Complete Task Analysis and Summary: Writ 6–8 B/C S602 Online

Table 2.1.3.4.2

Complete Task Analysis and Summary: Writ 6-8 B/C S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

2.1.3.5 Grades 9–12

Table 2.1.3.5.1

Complete Task Analysis and Summary: Writ 9–12 A S602 Online

Table 2.1.3.5.1

Complete Task Analysis and Summary: Writ 9-12 A S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

Table 2.1.3.5.2

Complete Task Analysis and Summary: Writ 9–12 B/C S602 Online

Table 2.1.3.5.2

Complete Task Analysis and Summary: Writ 9-12 B/C S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

2.1.4 Speaking

2.1.4.1 Grade 1

Table 2.1.4.1

Complete Task Analysis and Summary: Spek 1 S602 Online

Table 2.1.4.1
Complete Task Analysis and Summary: Spek 1 S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

2.1.4.2 Grades 2–3

Table 2.1.4.2

Complete Task Analysis and Summary: Spek 2-3 S602 Online

Table 2.1.4.2

Complete Task Analysis and Summary: Spek 2-3 S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

2.1.4.3 Grades 4–5

Table 2.1.4.3

Complete Task Analysis and Summary: Spek 4–5 S602 Online

Table 2.1.4.3
Complete Task Analysis and Summary: Spek 4–5 S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

2.1.4.4 Grades 6–8

Table 2.1.4.4

Complete Task Analysis and Summary: Spek 6-8 S602 Online

Table 2.1.4.4

Complete Task Analysis and Summary: Spek 6-8 S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

2.1.4.5 Grades 9–12

Table 2.1.4.5

Complete Task Analysis and Summary: Spek 9–12 S602 Online

Table 2.1.4.5

Complete Task Analysis and Summary: Spek 9-12 S602 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Information withheld due to confidentiality requirements.				

2.2 DIF Analysis and Summary

Before field testing, the Bias and Sensitivity Review Panel ensures that test items and tasks are free of material that (1) might favor any subgroup of students over another on the basis of gender, race/ethnicity, home language, religion, culture, region, or socioeconomic status, and (2) might be upsetting to students. This process is qualitatively driven, while the DIF analysis, described below, is data-driven. Please see Part 1, Section 2.3.1 for more information on Bias and Sensitivity panels.

CAL uses differential item functioning (DIF) analysis to investigate whether factors extraneous to English language proficiency (i.e., the construct being measured on the test) may have

influenced some students' performances on items. DIF attempts to find items that may be functioning differently for different groups based on criteria irrelevant to the construct that is purportedly being measured. We compare the performance of students on ACCESS for ELLs Online items and tasks by dividing students into two different groupings: first, males versus females; second, students of Hispanic ethnic background versus students of all other backgrounds. For the former analysis, females are the reference group, while males are the focal group. For the latter analysis, Hispanics are the reference group, while Non-Hispanics are the focal group. We exclude students for whom gender or ethnicity was unknown from both analyses. [Note: In the dataset, Hispanic ethnicity, as well as each of the race categories, is coded as a binary variable (Y/blank). Ethnicity information is counted as "Unknown" in cases where the student is recorded as blank for Hispanic ethnicity and also blank for every race category.] We used two commonly used procedures for detecting DIF: one for dichotomously scored items (Listening and Reading), conducted prior to operational testing, and one for polytomously scored items (Writing and Speaking), conducted on population data after the close of operational testing.

For dichotomous items, we used the Mantel-Haenszel (M-H) chi-square statistic (Mantel & Haenszel, 1959) procedure, originally proposed by the Educational Testing Service (ETS). This procedure compares item-level performances of students in the two groups (e.g., males versus females), who are divided into subgroups based on their performance on the total test. We assume that if there is no DIF, a similar percentage of students in each group should get the item correct at any ability level (based on performance on the total test). We use the M-H chi-square statistic to check the probability that the two groups performed comparably on each item across the ability groupings. The statistic is transformed into the "M-H delta" scale. This scale is symmetrical around zero, with a delta zero interpreted as indicating that neither group is favored. A positive result indicates that the focal group is favored; a negative result indicates that the reference group is favored.

The existing M-H procedure was designed for fixed forms, where all students take the same set of items; therefore, the students can be matched on the number-correct score when computing the M-H statistic. In the multistage computerized adaptive test condition, however, not all students take the same set of items; thus, it is not possible to match students on the number-correct score. Instead, we use a computerized adaptive test M-H DIF procedure (Zwick, Thayer, & Wingersky, 1993) to examine DIF for the Listening and Reading domains. First, we derive the student's expected true score for the entire item pool. To derive the expected true score, we transform each student's Rasch ability estimate into the expected true score metric by calculating the sum of the item response functions in the operational item pool, which is evaluated at the estimated ability level of the student. We use the expected true score of the students as the matching variable for the M-H DIF procedure. Once we have matched students on the expected true score, the ordinary M-H DIF procedure and the ETS evaluation criterion for severity of M-H DIF can be applied. In CAL's implementation of this method, students are matched for M-H DIF analysis based on this expected true score using two-unit intervals, as Zwick and Bridgeman (2014) recommended. We used a two-step purification process in conducting the DIF analysis; that is, we removed items with C-level DIF in the first

pass from the matching variable in the second stage, and then we recalculated the DIF for the remaining items.

Because DIF is measured on a continuous scale, and because most items are likely to show some degree of DIF, it is useful to have guidelines to determine when the level of DIF requires further review of the item. We follow the guidance provided by ETS (Zieky, 1993) to classify items into DIF levels as follows:

- A (no DIF) when the absolute value of delta is <1.0
- B (weak DIF) when the absolute value of delta is 1.0 to 1.5
- C (strong DIF) when the absolute value of delta is >1.5

For polytomous items (i.e., Writing and Speaking tasks), we took a similar approach. Our approach was based on the M-H chi-square statistic and the standardized mean difference following procedures that ETS developed (Allen, Carlson, & Zalanak, 1999; Zwick, Donoghue, & Grima, 1993). These DIF procedures for polytomous items were used to identify tasks that exhibit DIF. We used JMetrik (Meyer, 2018), an open-source computer program for psychometric analysis, to conduct the analyses. The procedures implemented in JMetrik first calculate the Cochran-Mantel-Haenszel chi-square statistic for testing statistical significance. This statistic gives an indication of the probability that observed differences are the result of chance but does not indicate how significant that difference is. To indicate how significant the difference is, we calculate the standardized mean difference between the performances of the two comparison groups. The standardized mean difference compares the means of the two groups, adjusting for differences in the distribution of the groups across the values of the total raw scores. To standardize the outcome, this difference is divided by the item score range and serves as an effect size measure for the Cochran-Mantel-Haenszel chi-square statistic. This effect size measure (reported as standardized P-DIF in JMetrik) ranges from -1 to 1 , which may present some interpretation challenges. To mitigate the negative value, the absolute value of the Cochran-Mantel-Haenszel chi-square statistic is used in JMetrik (Meyer, 2018) and the range of the rescaled effect size (standardized P-DIF*) is restricted to fall between 0 and 1 . The effect size flagging criterion for polytomous items that ETS proposed (Allen et al., 1999) is also rescaled to the standardized P-DIF* metric (Meyer, 2018).

Following guidance that ETS proposed for the National Assessment of Educational Progress (Allen et al., 1999), we classify ACCESS for ELLs Writing and Speaking tasks into three DIF levels as follows:

- AA (no DIF), when the Cochran-Mantel-Haenszel chi-square statistic is not significant or when it is significant and standardized P-DIF* is <0.05
- BB (weak DIF), when the Cochran-Mantel-Haenszel chi-square statistic is significant and standardized P-DIF* is ≥ 0.05 but <0.10
- CC (strong DIF) when the Cochran-Mantel-Haenszel chi-square statistic is significant and standardized P-DIF* is ≥ 0.10

The tables in this section provide a summary of the findings of the DIF analyses, by grade-level cluster, in the first table, followed by information for any item or task that showed B, BB, C, or CC-level DIF in the second table. The first column gives the DIF level: A, B, or C for dichotomous items or AA, BB, or CC for polytomous tasks (i.e., Writing and Speaking tasks).

The next columns show the contrasting groups in the DIF analyses: favoring male (focal group) versus favoring female (reference group) or favoring Hispanic (reference group) versus favoring non-Hispanic ethnicities (focal group). The top part of the table summarizes the number of items that exhibit DIF falling into each of the three categories (A, B, or C for Listening and Reading, and AA, BB, or CC for Writing and Speaking). Any items that show B (or BB) or C (or CC)–level DIF are reported in the second table.

If an item or task shows a C-level DIF, a DIF panel is convened. The DIF panel manager, from CAL, draws panelists from CAL staff members. Members are chosen so that a diverse background is represented. Therefore, the panel manager considers gender, first/second language backgrounds, and ethnicity when empaneling judges. The manager also ensures that some members have expertise in English as a Second Language instruction and/or professional development for teachers of ESL students. Without being told which items, if any, have an initial DIF finding, the panel is asked to discuss all items in the affected folder and come to a consensus on whether they believe or do not believe that the items demonstrate bias against a particular group and are or are not appropriate to place on the operational test.

For Listening and Reading items, we conduct DIF analysis and review prior to item selection, and we remove from the item selection pool any items that the panel judges to be inappropriate. Items that exhibited a C-level DIF but were judged to have no bias by the panel can be used in future series without the need to put the item before the panel again, per WIDA’s policy.

There is not sufficient scored data for DIF analysis of Speaking and Writing tasks prior to operational testing. We conduct DIF analysis using population data after operational testing is completed. Should a task exhibit CC-level DIF and should the review panel identify concern with that task, we recommend removal of the task from the subsequent year’s test.

For Series 602, one item in Listening grade 1 and one item in Listening grades 2–3 showed C-level DIF. These items were reviewed by a panel as described above, with both Listening grades 1 and 2–3 items being reviewed in previously held panels. These panels were not able to detect any reason for bias in the performance of these items and recommended that the items be retained on the assessment.

2.2.1 Listening

2.2.1.1 Grade 1

Table 2.2.1.1.1

DIF Analysis over Gender and Ethnicities: List 1 S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.1.1.2

DIF Summary: List 1 S602 Online

Task/Item Name	DIF Level (F/M)	Favored Group (F/M)	DIF Level (H/O)	Favored Group (H/O)
Information withheld due to confidentiality requirements.				

2.2.1.2 Grades 2–3

Table 2.2.1.2.1

DIF Analysis over Gender and Ethnicities: List 2–3 S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.1.2.2

DIF Summary: List 2–3 S602 Online

Task/Item Name	DIF Level (F/M)	Favored Group (F/M)	DIF Level (H/O)	Favored Group (H/O)
Information withheld due to confidentiality requirements.				

2.2.1.3 Grades 4–5

Table 2.2.1.3.1

DIF Analysis over Gender and Ethnicities: List 4–5 S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.1.3.2

DIF Summary: List 4-5 S602 Online

Task/Item Name	DIF Level (F/M)	Favored Group (F/M)	DIF Level (H/O)	Favored Group (H/O)
Information withheld due to confidentiality requirements.				

2.2.1.4 Grades 6–8

Table 2.2.1.4

DIF Analysis and Summary: List 6-8 S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.1.5 Grades 9–12

Table 2.2.1.5.1

DIF Analysis over Gender and Ethnicities: List 9-12 S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.1.5.2

DIF Summary: List 9-12 S602 Online

Task/Item Name	DIF Level (F/M)	Favored Group (F/M)	DIF Level (H/O)	Favored Group (H/O)
1.	Information withheld due to confidentiality requirements.			
7				
1				
2				

2.2.2 Reading

2.2.2.1 Grade 1

Table 2.2.2.1.1

DIF Analysis over Gender and Ethnicities: Read 1 S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.2.1.2

DIF Summary: Read 1 S602 Online

Task/Item Name	DIF Level (F/M)	Favored Group (F/M)	DIF Level (H/O)	Favored Group (H/O)
Information withheld due to confidentiality requirements.				

2.2.2.2 Grades 2–3

Table 2.2.2.2

DIF Analysis and Summary: Read 2-3 S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.2.3 Grades 4–5

Table 2.2.2.3.1

DIF Analysis over Gender and Ethnicities: Read 4-5 S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.2.3.2

DIF Summary: Read 4-5 S602 Online

Task/Item Name	DIF Level (F/M)	Favored Group (F/M)	DIF Level (H/O)	Favored Group (H/O)
Information withheld due to confidentiality requirements.				

2.2.2.4 Grades 6–8

Table 2.2.2.4.1

DIF Analysis over Gender and Ethnicities: Read 6-8 S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.2.4.2

DIF Summary: Read 6-8 S602 Online

Task/Item Name	DIF Level (F/M)	Favored Group (F/M)	DIF Level (H/O)	Favored Group (H/O)
Information withheld due to confidentiality requirements.				

2.2.2.5 Grades 9–12

Table 2.2.2.5.1

DIF Analysis over Gender and Ethnicities: Read 9-12 S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.2.5.2

DIF Summary: Read 9-12 S602 Online

Task/Item Name	DIF Level (F/M)	Favored Group (F/M)	DIF Level (H/O)	Favored Group (H/O)
Information withheld due to confidentiality requirements.				

2.2.3 Writing

2.2.3.1 Grade 1

Table 2.2.3.1.1

DIF Analysis and Summary: Writ 1 A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.3.1.2

DIF Analysis and Summary: Writ 1 B/C S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.3.2 Grades 2-3

Table 2.2.3.2.1

DIF Analysis and Summary: Writ 2-3 A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.3.2.2

DIF Analysis and Summary: Writ 2-3 B/C S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.3.3 Grades 4–5

Table 2.2.3.3.1

DIF Analysis and Summary: Writ 4-5 A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.3.3.2

DIF Analysis and Summary: Writ 4-5 B/C S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.3.4 Grades 6–8

Table 2.2.3.4.1

DIF Analysis and Summary: Writ 6-8 A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.3.4.2

DIF Analysis and Summary: Writ 6-8 B/C S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.3.5 Grades 9–12

Table 2.2.3.5.1

DIF Analysis and Summary: Writ 9-12 A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.3.5.2

DIF Analysis and Summary: Writ 9-12 B/C S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.4 Speaking

2.2.4.1 Grade 1

Table 2.2.4.1.1

DIF Analysis and Summary: Spek 1 Pre-A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.4.1.2

DIF Analysis and Summary: Spek 1 A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.4.1.3

DIF Analysis and Summary: Spek 1 B/C S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.4.2 Grades 2–3

Table 2.2.4.2.1

DIF Analysis and Summary: Spek 2-3 Pre-A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.4.2.2

DIF Analysis and Summary: Spek 2-3 A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.4.2.3

DIF Analysis and Summary: Spek 2-3 B/C S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.4.3 Grades 4–5

Table 2.2.4.3.1

DIF Analysis and Summary: Spek 4-5 Pre-A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.4.3.2

DIF Analysis and Summary: Spek 4-5 A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.4.3.3

DIF Analysis and Summary: Spek 4-5 B/C S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.4.4 Grades 6–8

Table 2.2.4.4.1

DIF Analysis and Summary: Spek 6-8 Pre-A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.4.4.2

DIF Analysis and Summary: Spek 6-8 A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.4.4.3

DIF Analysis and Summary: Spek 6-8 B/C S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.2.4.5 Grades 9–12

Table 2.2.4.5.1

DIF Analysis and Summary: Spek 9-12 Pre-A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.4.5.2

DIF Analysis and Summary: Spek 9-12 A S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

Table 2.2.4.5.3

DIF Analysis and Summary: Spek 9-12 B/C S602 Online

DIF Level	Favoring Male	Favoring Female	Favoring Hispanic	Favoring Others
Information withheld due to confidentiality requirements.				

2.3 *Raw Score Distribution*

Figures and tables in this section provide raw score information for Speaking and Writing. For each grade-level cluster and tier combination, the figure shows the distribution of the raw scores. The horizontal axis shows the raw scores. The vertical axis shows the number of students (count). Each bar shows how many students received each raw score.

Each table in this section summarizes results for a grade-level cluster and tier combination (e.g., Speaking 4–5 Tier A). For each table, results are broken down by grade and presented for the grade-level cluster for that tier. The following information is included in each table:

- The number of students in the analyses (the number of students who were not absent, invalid, refused, exempt, or in the wrong grade-level cluster)
- The minimum observed raw score
- The maximum observed raw score
- The mean (average) raw score
- The standard deviation (std. dev.) of the raw scores

Test design and student population impact the distribution of raw scores. In general, raw score distributions tend to be smoothly distributed with a single peak; however, there are several exceptions. Understanding these distributions supports the understanding of other statistical properties of the test forms.

Speaking Pre-A forms are designed for students at the very earliest stages of English language proficiency. Students routed to the Pre-A form have very low performances on Listening and Reading and are administered three Speaking tasks, each scored 0 to 2, for a total raw score range of 0 to 6. Tasks on the Pre-A form are, by design, very easy and intended to ensure beginning students are not discouraged. Large numbers of students can achieve all six points on this form. Students routed to the A form take three PL 1 tasks, scored 0 to 2, and three PL 3 tasks, scored 0 to 4, for a total raw score range of 0 to 18. Students routed to take the B/C form did not take the P1 tasks. These students take three PL 3 and three PL 5 tasks, each scoring 0 to 4. The total raw score range for the Tier B/C form is 0 to 24. Note that this is a scoring change for the Series 602 test as in the past we awarded these students two points on each of the three P1 tasks.

2.3.1 **Listening**

The ACCESS 2.0 Online Listening test is a multistage adaptive assessment. As students do not all take the same set of items in the test, raw score distributions are not presented.

2.3.2 **Reading**

The ACCESS 2.0 Online Reading test is a multistage adaptive assessment. As students do not all take the same set of items in the test, raw score distributions are not presented.

2.3.3 Writing

2.3.3.1 Grade 1

Table 2.3.3.1.1

Raw Score Descriptive Statistics: Writ 1 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	209,593	0	14	5.18	3.08
Total	209,593	0	14	5.18	3.08

Figure 2.3.3.1.1

Raw Scores: Writ A S602 Online

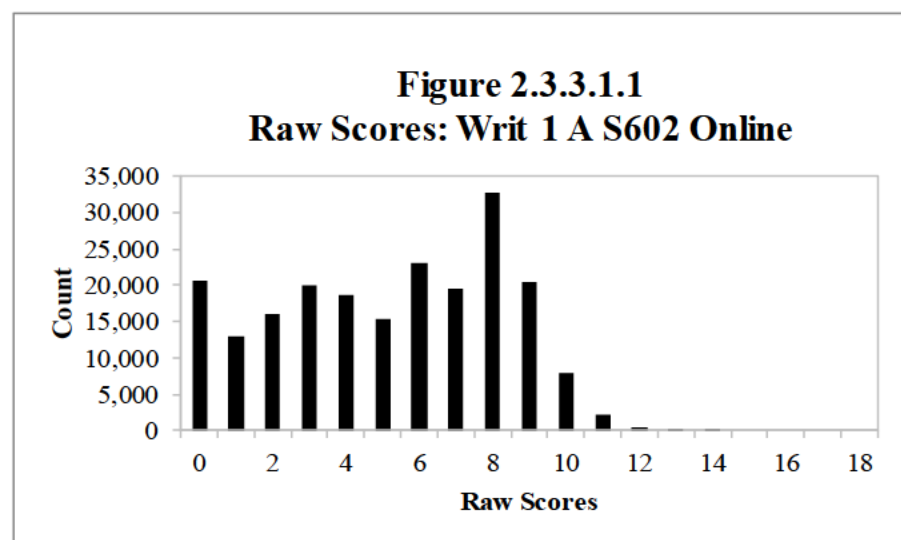


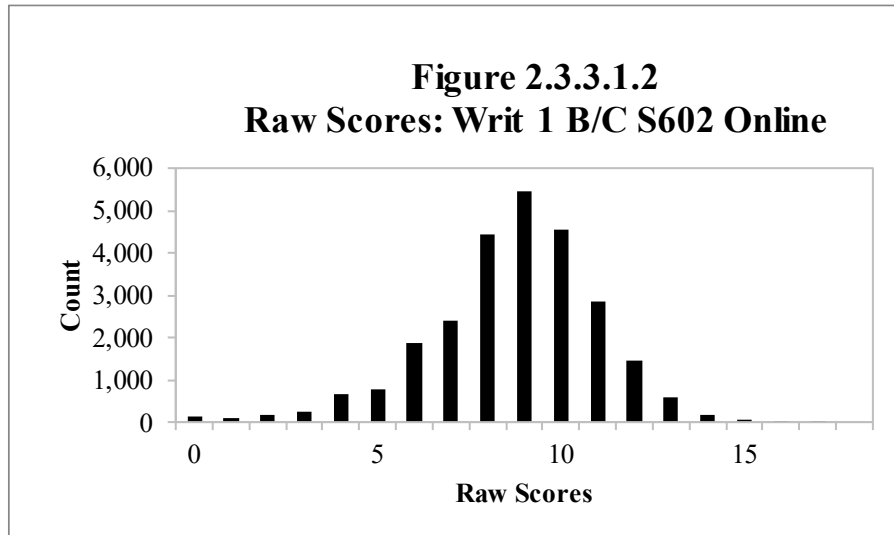
Table 2.3.3.1.2

Raw Score Descriptive Statistics: Writ 1 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	26,141	0	17	8.70	2.33
Total	26,141	0	17	8.70	2.33

Figure 2.3.3.1.2

Raw Scores: Writ 1 B/C S602 Online



2.3.3.2 Grades 2–3

Table 2.3.3.2.1

Raw Score Descriptive Statistics: Writ 2–3 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	87,830	0	15	4.79	3.40
3	73,330	0	16	5.44	3.56
Total	161,160	0	16	5.08	3.49

Figure 2.3.3.2.1

Raw Scores: Writ 2–3 A S602 Online

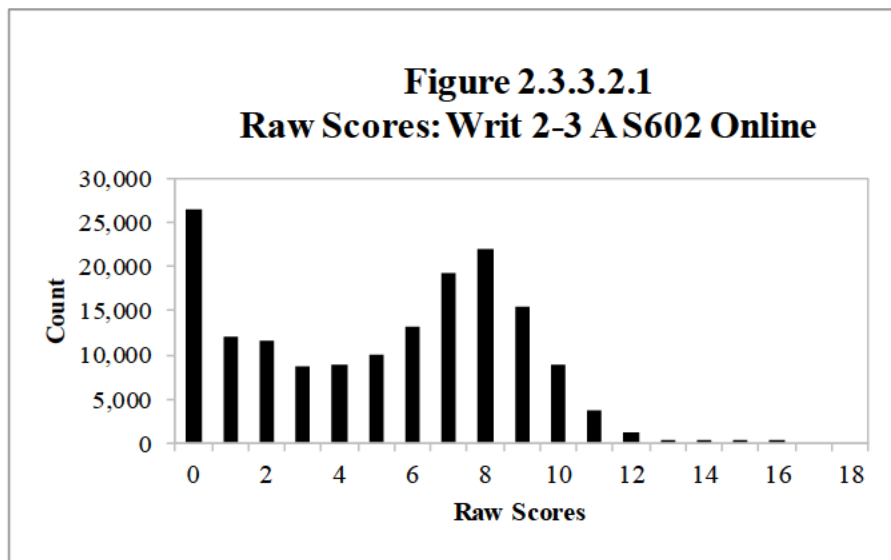


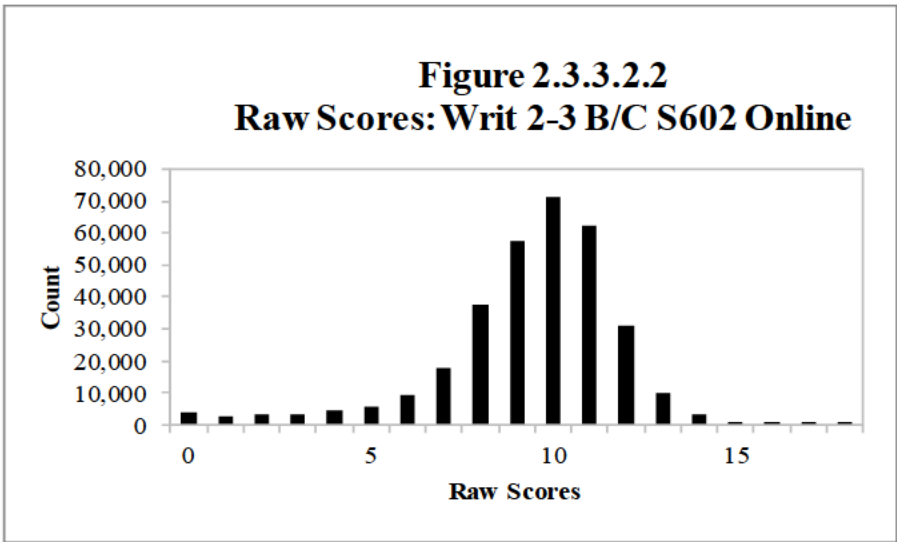
Table 2.3.3.2.2

Raw Score Descriptive Statistics: Writ 2–3 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	159,427	0	18	8.61	2.66
3	164,235	0	18	10.01	2.09
Total	323,662	0	18	9.32	2.48

Figure 2.3.3.2.2

Raw Scores: Writ 2–3 B/C S602 Online



2.3.3.3 Grades 4–5

Table 2.3.3.3.1

Raw Score Descriptive Statistics: Writ 4–5 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	56,044	0	15	4.09	3.13
5	53,170	0	14	4.76	3.28
Total	109,214	0	15	4.41	3.22

Figure 2.3.3.3.1

Raw Scores: Writ 4–5 A S602 Online

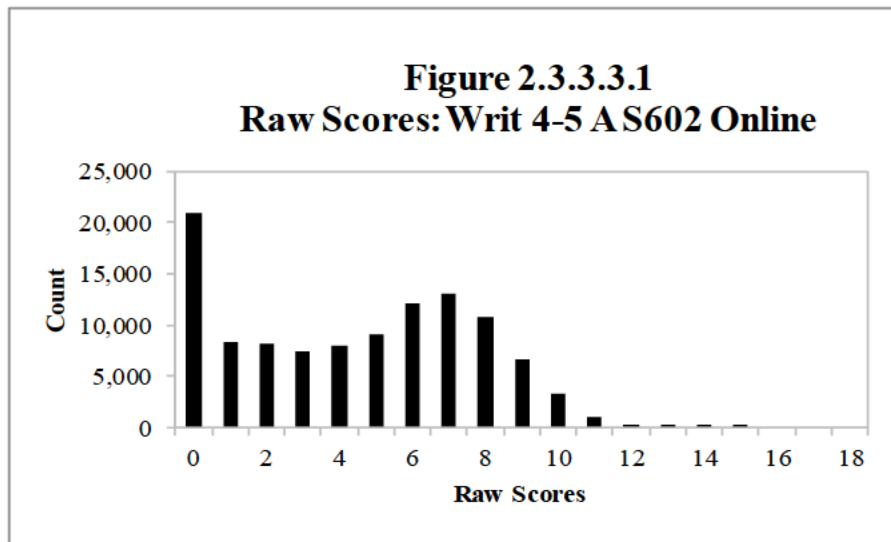


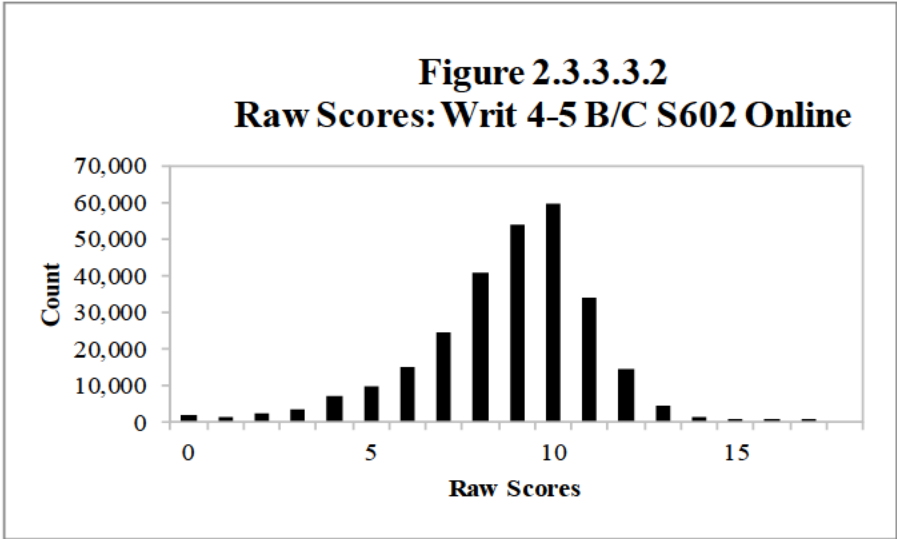
Table 2.3.3.3.2

Raw Score Descriptive Statistics: Writ 4–5 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	151,995	0	17	8.36	2.42
5	120,478	0	17	9.17	2.10
Total	272,473	0	17	8.72	2.32

Figure 2.3.3.3.2

Raw Scores: Writ 4–5 B/C S602 Online



2.3.3.4 Grades 6–8

Table 2.3.3.4.1

Raw Score Descriptive Statistics: Writ 6–8 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	61,057	0	15	5.58	3.05
7	72,746	0	16	6.39	3.09
8	73,194	0	17	6.84	3.13
Total	206,997	0	17	6.31	3.13

Figure 2.3.3.4.1

Raw Scores: Writ 6–8 A S602 Online

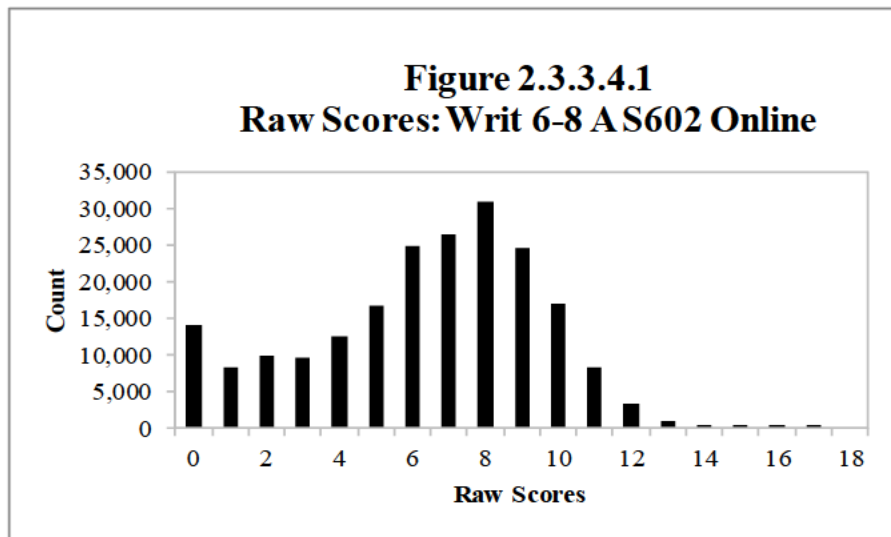


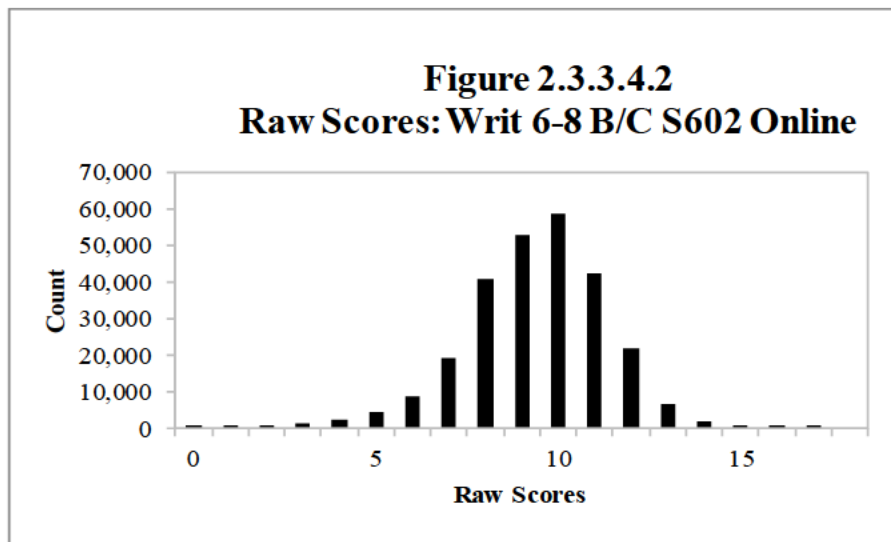
Table 2.3.3.4.2

Raw Score Descriptive Statistics: Writ 6–8 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	91,844	0	16	8.53	2.04
7	86,756	0	16	9.46	1.83
8	83,536	0	17	10.07	1.78
Total	262,136	0	17	9.33	2.00

Figure 2.3.3.4.1

Raw Scores: Writ 6–8 B/C S602 Online



2.3.3.5 Grades 9–12

Table 2.3.3.5.1

Raw Score Descriptive Statistics: Writ 9–12 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	70,048	0	17	5.72	3.52
10	52,662	0	16	6.35	3.26
11	41,969	0	16	6.91	3.14
12	27,325	0	17	7.14	3.08
Total	192,004	0	17	6.36	3.35

Figure 2.3.3.5.1

Raw Scores: Writ 9–12 A S602 Online

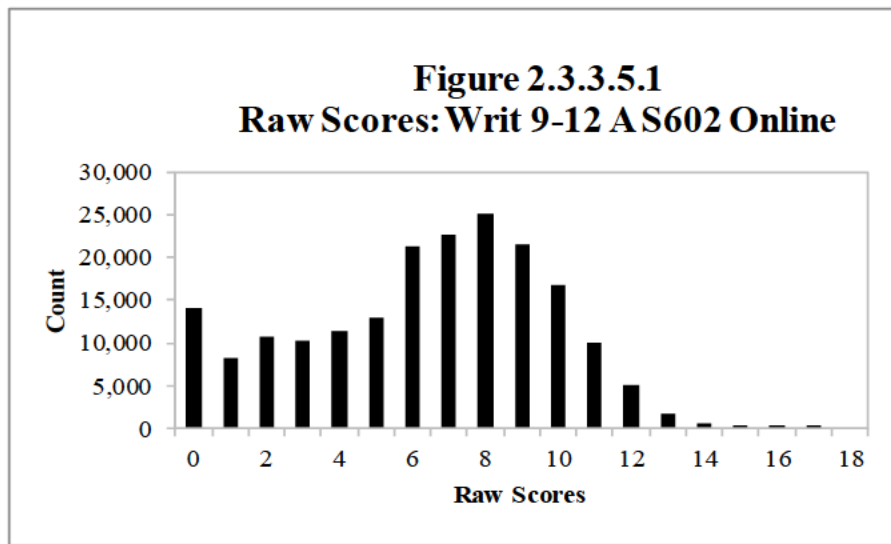


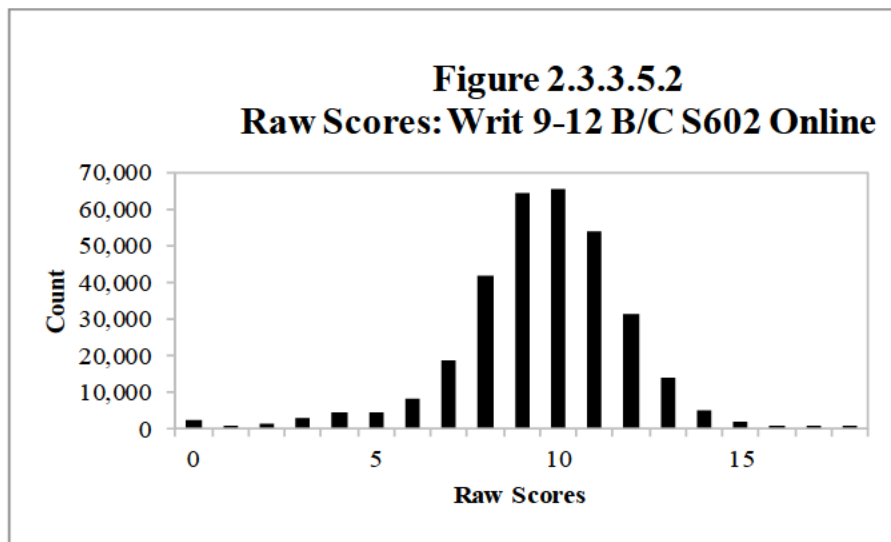
Table 2.3.3.5.2

Raw Score Descriptive Statistics: Writ 9–12 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	98,574	0	17	9.31	2.35
10	91,888	0	18	9.50	2.28
11	73,351	0	17	9.74	2.18
12	56,123	0	17	9.62	2.24
Total	319,936	0	18	9.52	2.28

Figure 2.3.3.5.2

Raw Scores: Writ 9–12 B/C S602 Online



2.3.4 Speaking

2.3.4.1 Grade 1

Table 2.3.4.1.1

Raw Score Descriptive Statistics: Spek 1 Pre-A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	14,805	0	6	3.95	2.27
Total	14,805	0	6	3.95	2.27

Figure 2.3.4.1.1

Raw Scores: Spek 1 Pre-A S602 Online

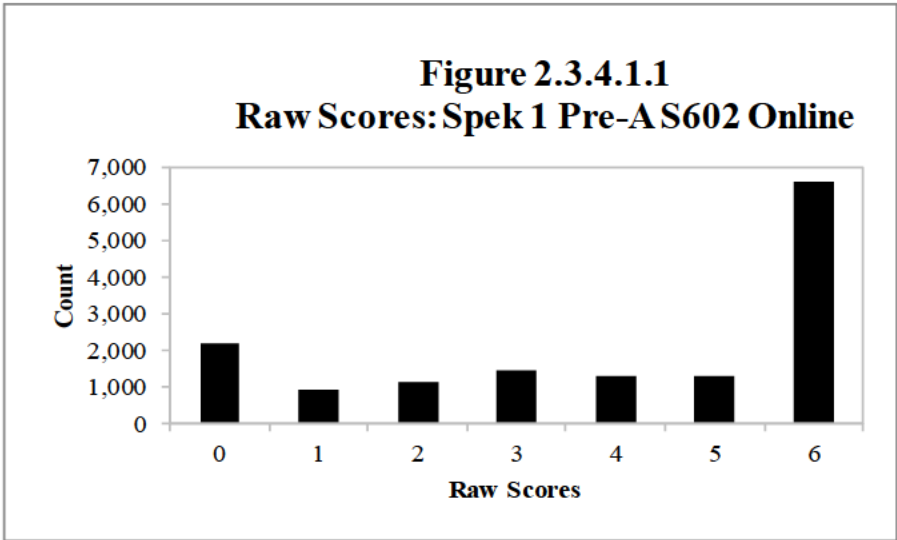


Table 2.3.4.1.2

Raw Score Descriptive Statistics: Spek 1 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	108,591	0	18	10.03	3.98
Total	108,591	0	18	10.03	3.98

Figure 2.3.4.1.2

Raw Scores: Spek 1 A S602 Online

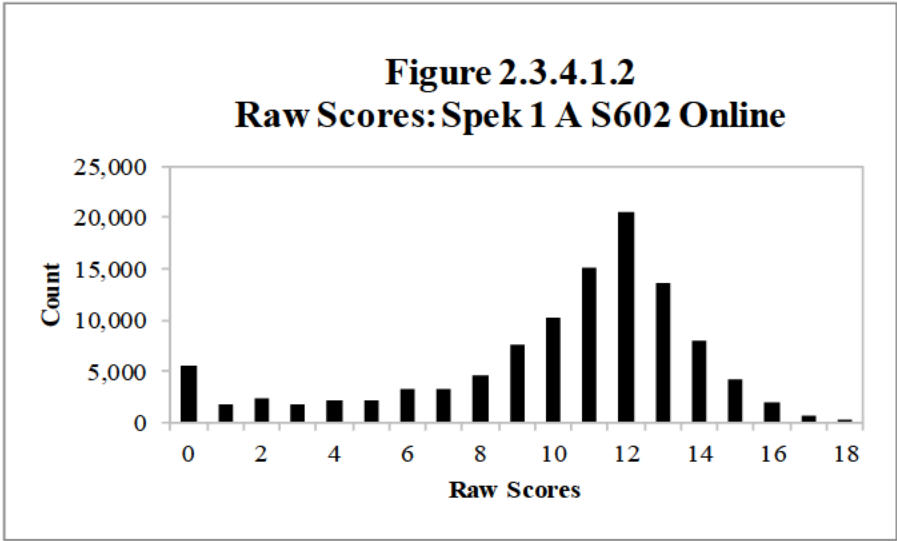


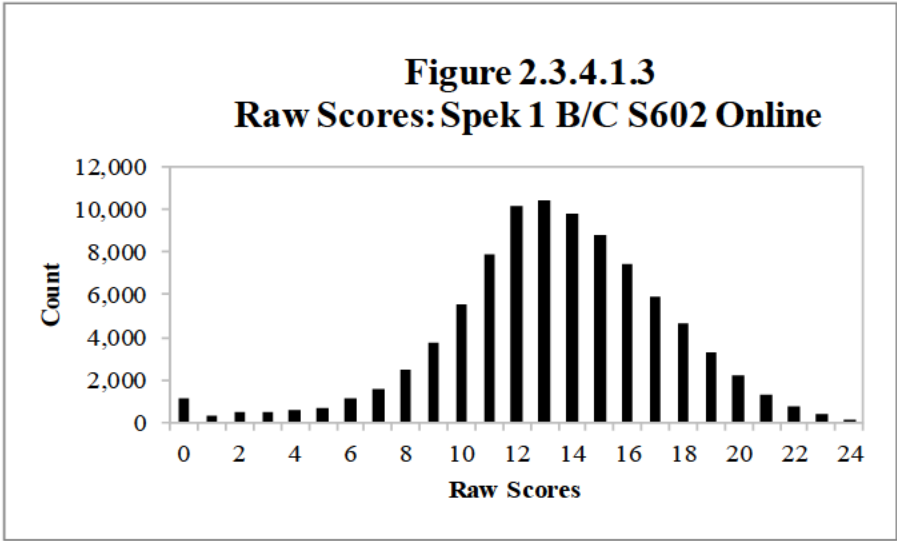
Table 2.3.4.1.3

Raw Score Descriptive Statistics: Spek 1 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	91,009	0	24	13.29	4.05
Total	91,009	0	24	13.29	4.05

Figure 2.3.4.1.3

Raw Scores: Spek 1 B/C S602 Online



2.3.4.2 Grades 2–3

Table 2.3.4.2.1

Raw Score Descriptive Statistics: Spek 2–3 Pre-A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	9,393	0	6	4.22	2.25
3	20,574	0	6	4.34	2.21
Total	29,967	0	6	4.30	2.22

Figure 2.3.4.2.1

Raw Scores: Spek 2–3 Pre-A S602 Online

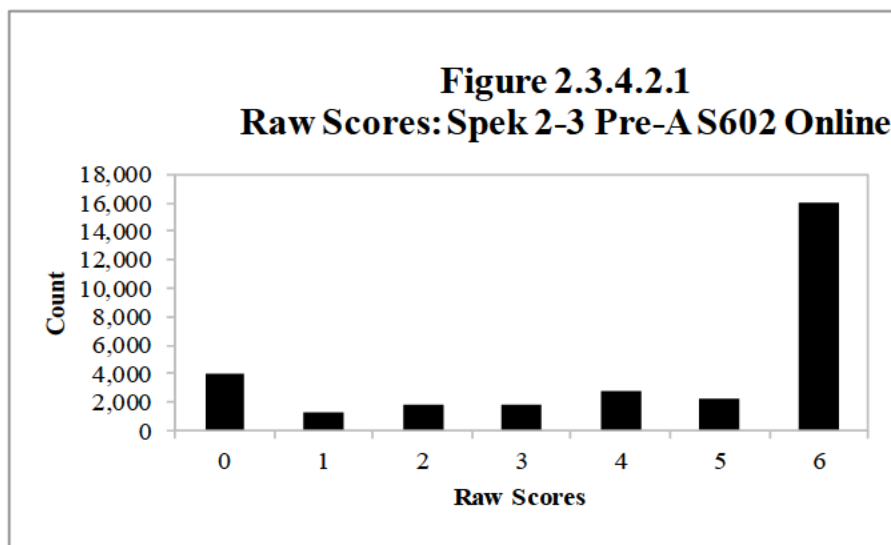


Table 2.3.4.2.2

Raw Score Descriptive Statistics: Spek 2-3 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	74,659	0	18	9.69	3.93
3	67,665	0	18	11.15	3.51
Total	142,324	0	18	10.38	3.80

Figure 2.3.4.2.2

Raw Scores: Spek 2-3 A S602 Online

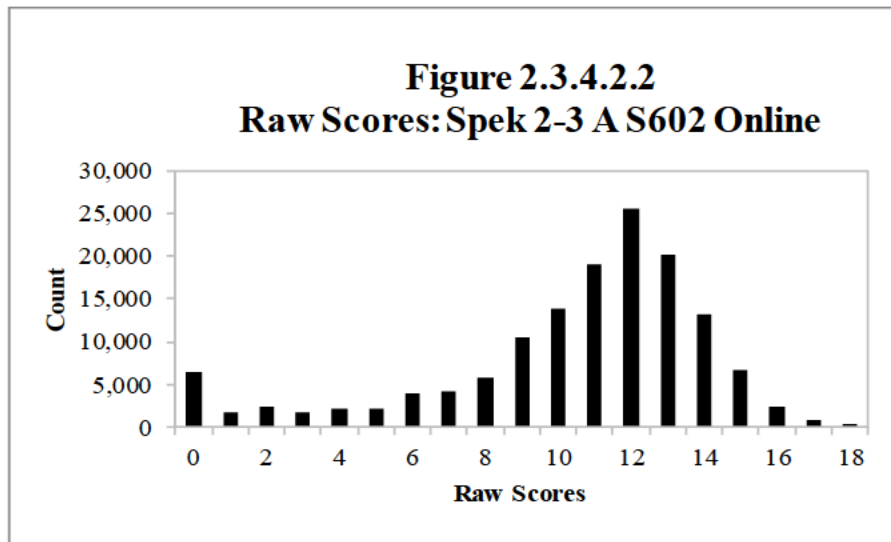


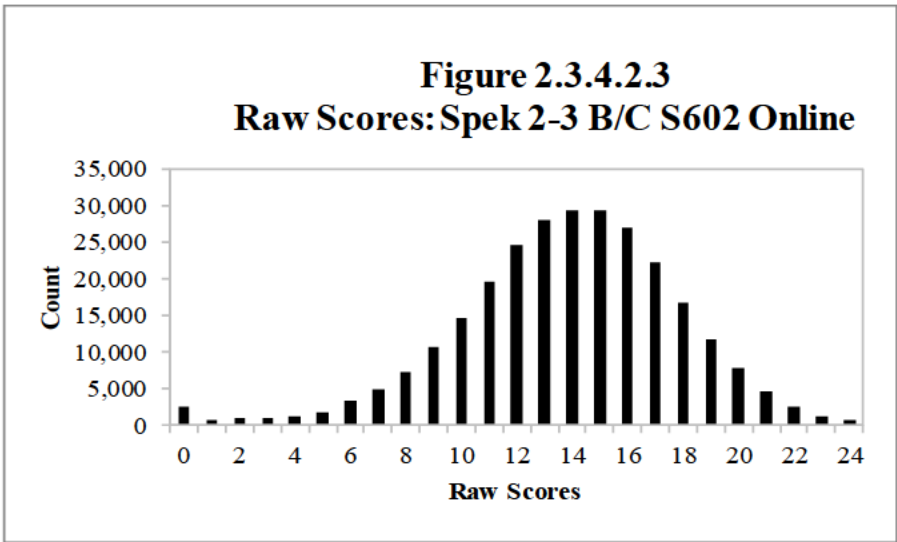
Table 2.3.4.2.3

Raw Score Descriptive Statistics: Spek 2–3 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	141,555	0	24	12.91	3.98
3	131,745	0	24	14.76	3.72
Total	273,300	0	24	13.80	3.97

Figure 2.3.4.2.3

Raw Scores: Spek 2–3 B/C S602 Online



2.3.4.3 Grades 4–5

Table 2.3.4.3.1

Raw Score Descriptive Statistics: Spek 4–5 Pre-A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	4,046	0	6	3.75	2.25
5	8,610	0	6	4.01	2.20
Total	12,656	0	6	3.92	2.22

Figure 2.3.4.3.1

Raw Scores: Spek 4–5 Pre-A S602 Online

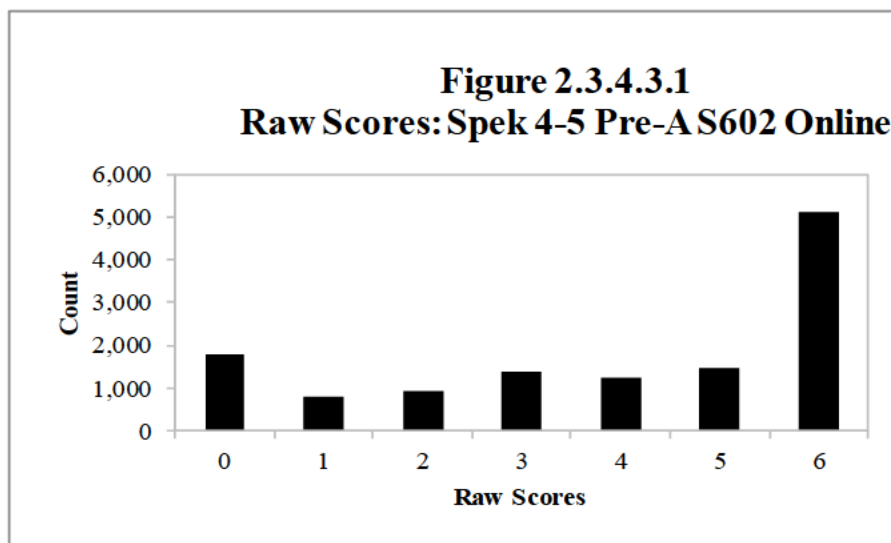


Table 2.3.4.3.2

Raw Score Descriptive Statistics: Spek 4–5 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	43,530	0	18	8.85	3.59
5	32,772	0	18	9.11	3.51
Total	76,302	0	18	8.96	3.56

Figure 2.3.4.3.2

Raw Scores: Spek 4–5 A S602 Online

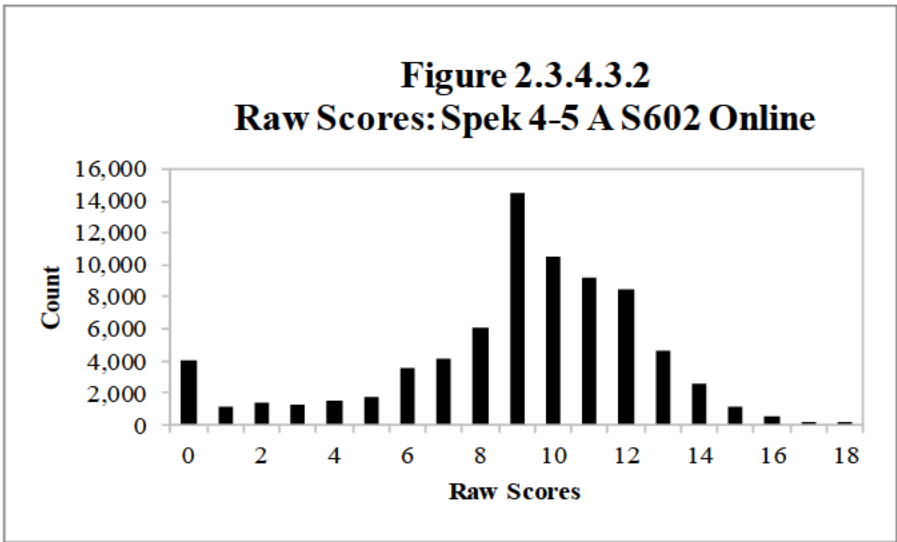


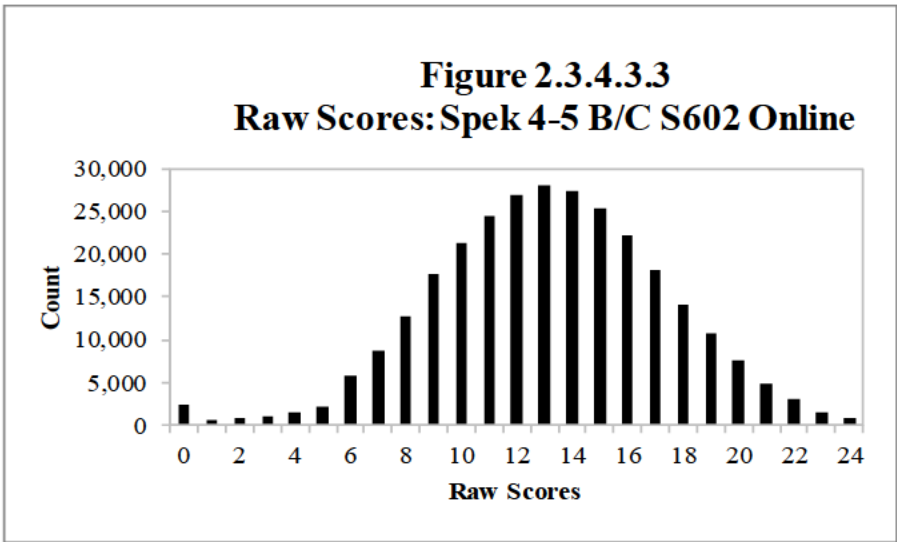
Table 2.3.4.3.3

Raw Score Descriptive Statistics: Spek 4–5 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	158,569	0	24	13.09	4.12
5	130,202	0	24	13.21	4.14
Total	288,771	0	24	13.14	4.13

Figure 2.3.4.3.3

Raw Scores: Spek 4–5 B/C S602 Online



2.3.4.4 Grades 6–8

Table 2.3.4.4.1

Raw Score Descriptive Statistics: Spek 6–8 Pre-A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	6,152	0	6	4.36	2.18
7	10,204	0	6	4.48	2.13
8	14,536	0	6	4.61	2.08
Total	30,892	0	6	4.52	2.12

Figure 2.3.4.4.1

Raw Scores: Spek 6–8 Pre-A S602 Online

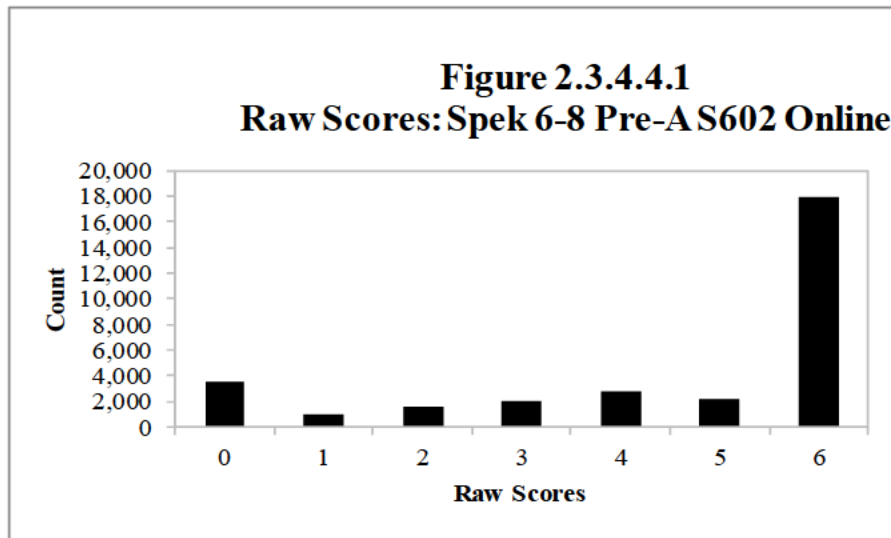


Table 2.3.4.4.2

Raw Score Descriptive Statistics: Spek 6–8 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	35,089	0	18	9.15	3.62
7	25,694	0	18	9.02	3.60
8	47,042	0	18	10.49	3.54
Total	107,825	0	18	9.70	3.64

Figure 2.3.4.4.2

Raw Scores: Spek 6–8 A S602 Online

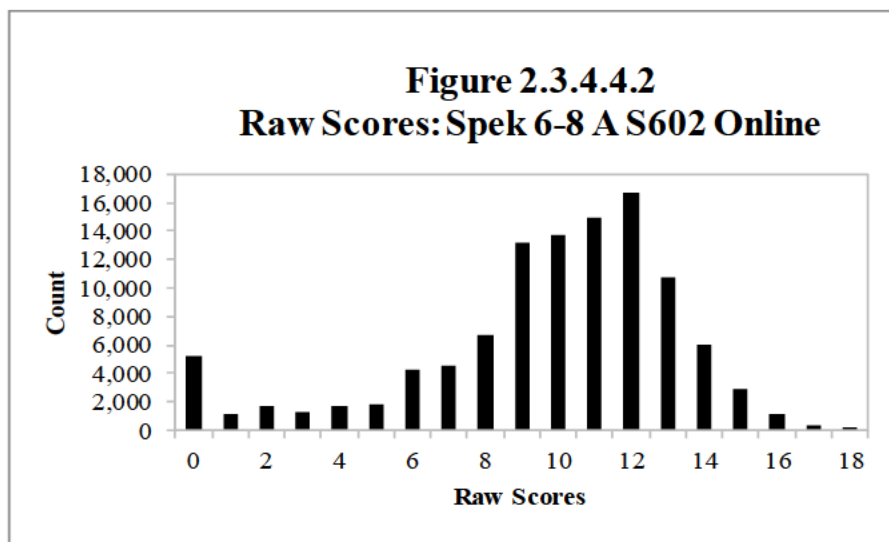


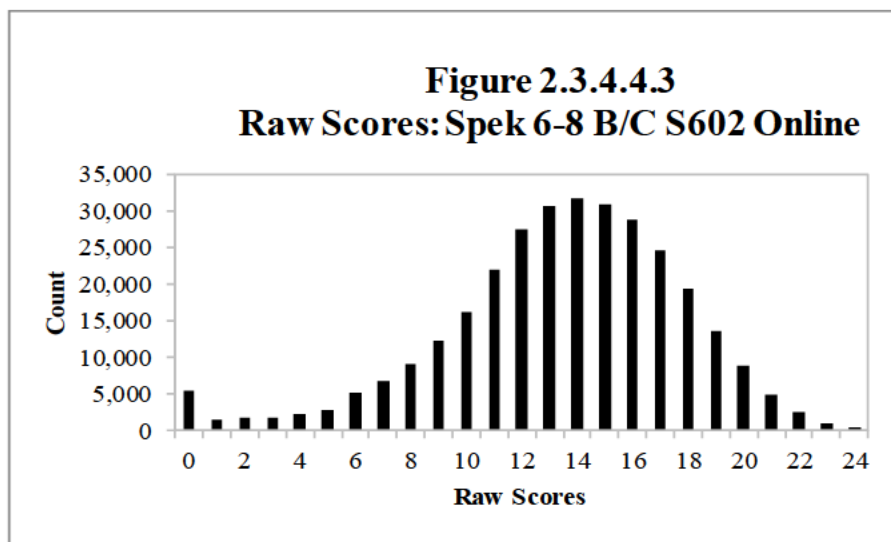
Table 2.3.4.4.3

Raw Score Descriptive Statistics: Spek 6–8 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	104,526	0	24	12.85	4.08
7	115,424	0	24	13.17	4.40
8	89,410	0	24	14.53	4.21
Total	309,360	0	24	13.46	4.30

Figure 2.3.4.4.2

Raw Scores: Spek 6–8 B/C S602 Online



2.3.4.5 Grades 9–12

Table 2.3.4.5.1

Raw Score Descriptive Statistics: Spek 9–12 Pre-A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	9,161	0	6	4.48	2.06
10	9,342	0	6	4.92	1.86
11	11,193	0	6	5.04	1.85
12	8,498	0	6	5.09	1.86
Total	38,194	0	6	4.89	1.92

Figure 2.3.4.5.1

Raw Scores: Spek 9–12 Pre-A S602 Online

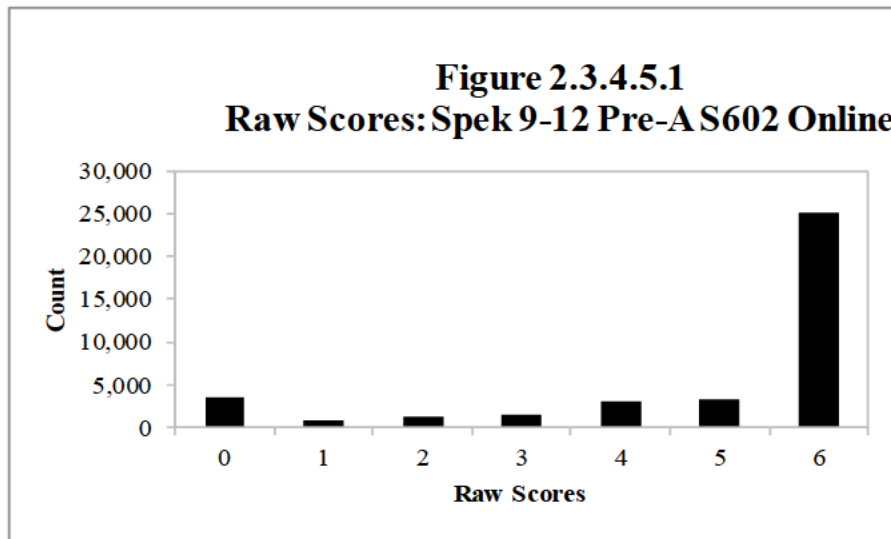


Table 2.3.4.5.2

Raw Score Descriptive Statistics: Spek 9–12 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	94,373	0	18	9.96	3.68
10	60,308	0	18	10.13	3.57
11	24,935	0	18	10.00	3.51
12	39,411	0	18	11.19	3.54
Total	219,027	0	18	10.23	3.63

Figure 2.3.4.5.2

Raw Scores: Spek 9–12 A S602 Online

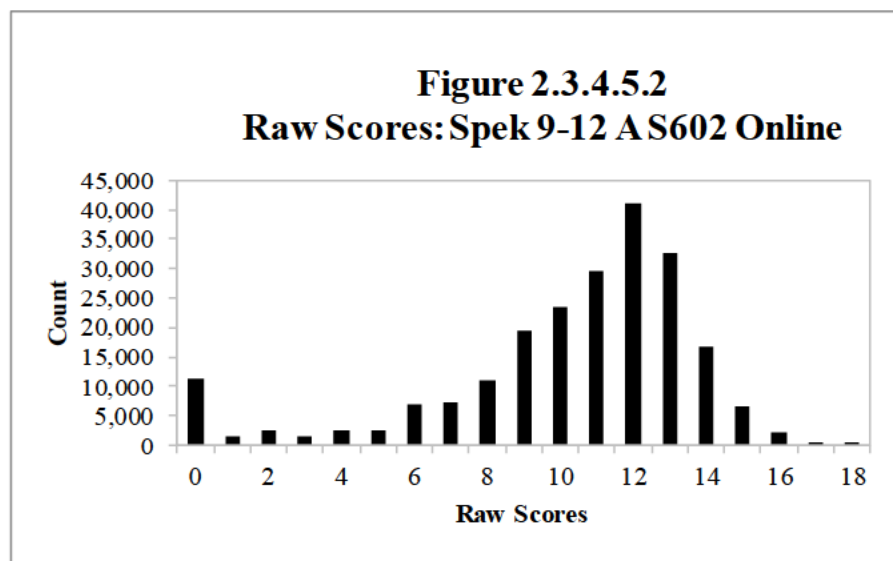


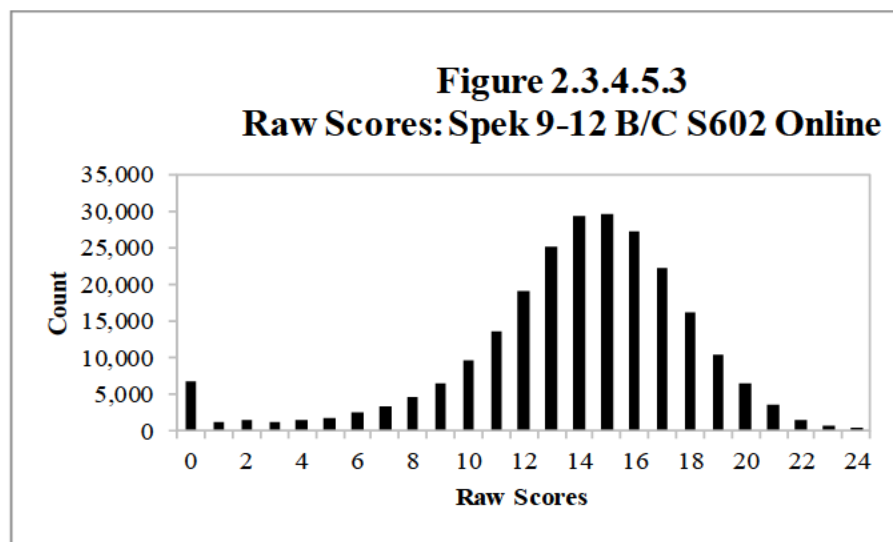
Table 2.3.4.5.3

Raw Score Descriptive Statistics: Spek 9–12 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	61,928	0	24	13.71	4.04
10	71,692	0	24	13.77	4.24
11	75,932	0	24	13.44	4.55
12	34,390	0	24	14.46	4.27
Total	243,942	0	24	13.75	4.31

Figure 2.3.4.5.3

Raw Scores: Spek 9–12 B/C S602 Online



2.4 Scale Score Distribution

The figures and tables in this section relate to the ACCESS for ELLs scale scores on each test form. We converted raw scores to vertically equated scale scores for each test form. The scale score distributions are presented by grade-level cluster. Additionally, for Writing and Speaking, we present the distributions by grade-level cluster and tier.

For each test form, the figure shows the distribution of the scale scores. Scale scores are plotted on the horizontal axis.

For Listening and Reading, we grouped the scale scores into units of five scale score points (e.g., 100–104, 105–109, 110–114, etc.). It should be noted that the scale score distribution is presented by grade level cluster. Because the Listening and Reading domains are computer adaptive, students were routed by the engine into one of three different tier folders across stages, where the folders differ in difficulties. Therefore, in some plots in this section, it may appear that there is more than one set of data presented.

For Speaking and Writing, we plotted each individual scale score point for each test form. For figures that summarize both test forms in a cluster, we grouped scale scores into units of five scale score points.

It should be noted that Speaking Pre-A forms are designed for students at the very earliest stages of English language proficiency. Students routed to the Pre-A form have very low performances on Listening and Reading and are administered three Speaking tasks, each scored 0 to 2, for a total raw score range of 0 to 6. Tasks on the Pre-A form are by design very easy and intended to ensure beginning students are not discouraged. Therefore, large numbers of students can achieve all 6 points on this form as reflected in the Pre-A tables and figures in this section.

The number of students with scale scores falling into each range is plotted on the vertical axis.

The tables in this section show, by grade and by total for the grade-level cluster:

- The number of students in the analyses (count)
- The minimum observed scale score
- The maximum observed scale score
- The mean (average) scale score
- The standard deviation (std. dev.) of the scale score

2.4.1 Listening

2.4.1.1 Grade 1

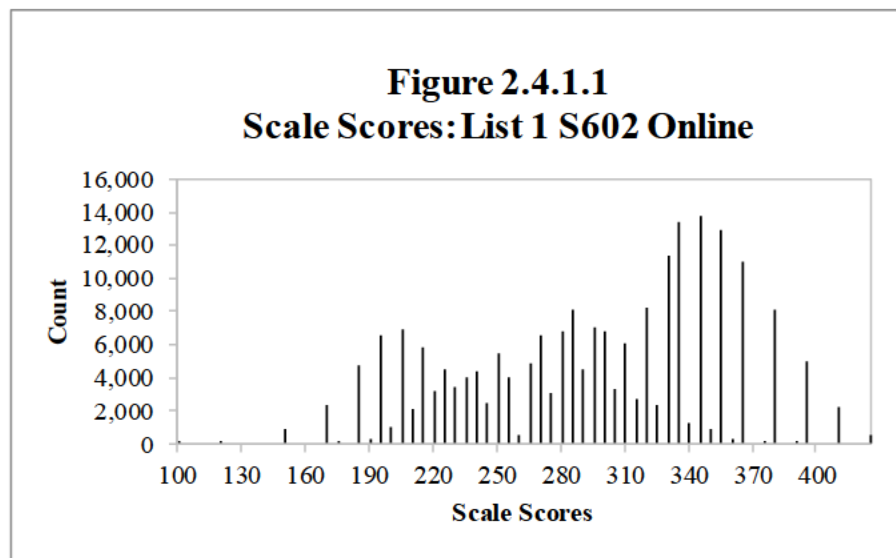
Table 2.4.1.1

Scale Score Descriptive Statistics: List 1 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	213,555	104	425	296.86	59.37
Total	213,555	104	425	296.86	59.37

Figure 2.4.1.1.

Scale Scores: List 1 S602 Online



2.4.1.2 Grades 2–3

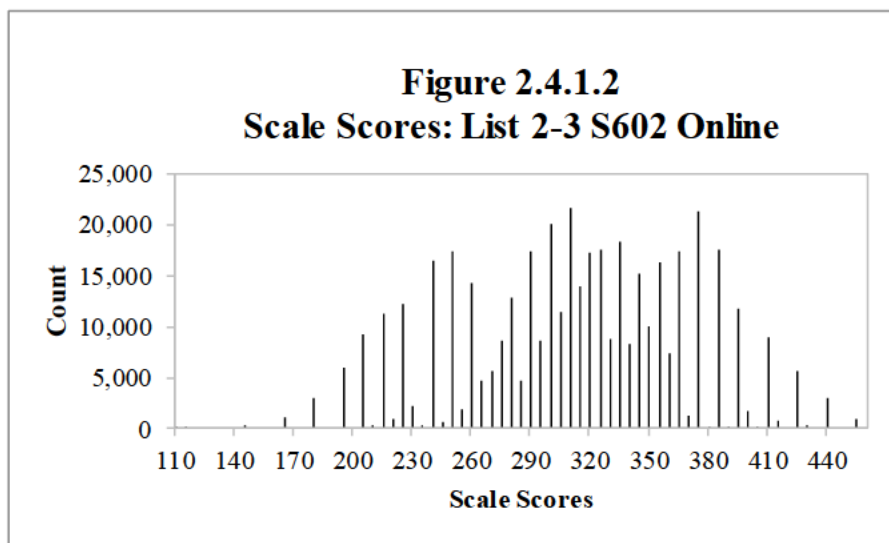
Table 2.4.1.2

Scale Score Descriptive Statistics: List 2–3 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	221,182	112	457	303.79	53.56
3	215,746	112	457	323.92	58.53
Total	436,928	112	457	313.73	56.96

Figure 2.4.1.2

Scale Scores: List 2–3 S602 Online



2.4.1.3 Grades 4–5

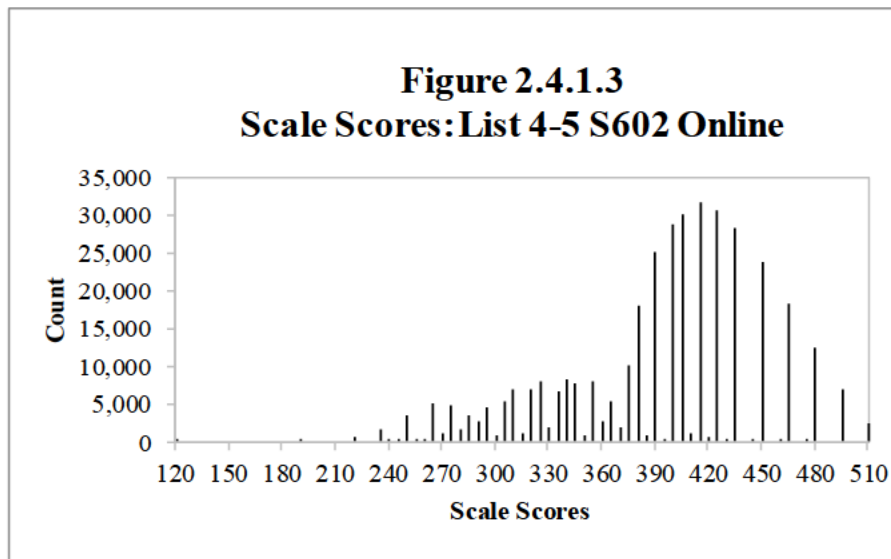
Table 2.4.1.3

Scale Score Descriptive Statistics: List 4–5 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	203,631	120	511	393.00	55.11
5	169,686	120	511	397.25	59.18
Total	373,317	120	511	394.93	57.04

Figure 2.4.1.3

Scale Scores: List 4–5 S602 Online



2.4.1.4 Grades 6–8

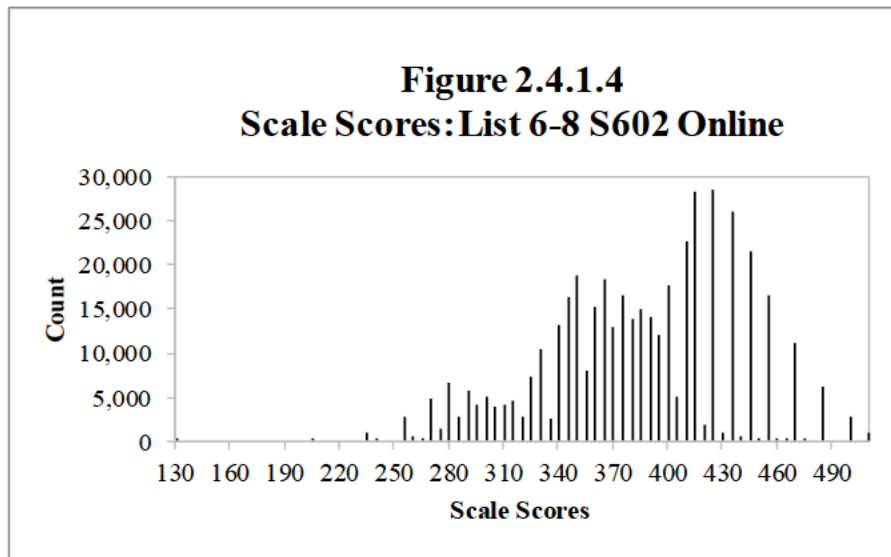
Table 2.4.1.4

Scale Score Descriptive Statistics: List 6–8 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	140,833	132	514	379.43	46.25
7	147,693	132	514	387.12	51.09
8	146,994	132	514	392.00	55.43
Total	435,520	132	514	386.28	51.38

Figure 2.4.1.4

Scale Scores: List 6–8 S602 Online



2.4.1.5 Grades 9–12

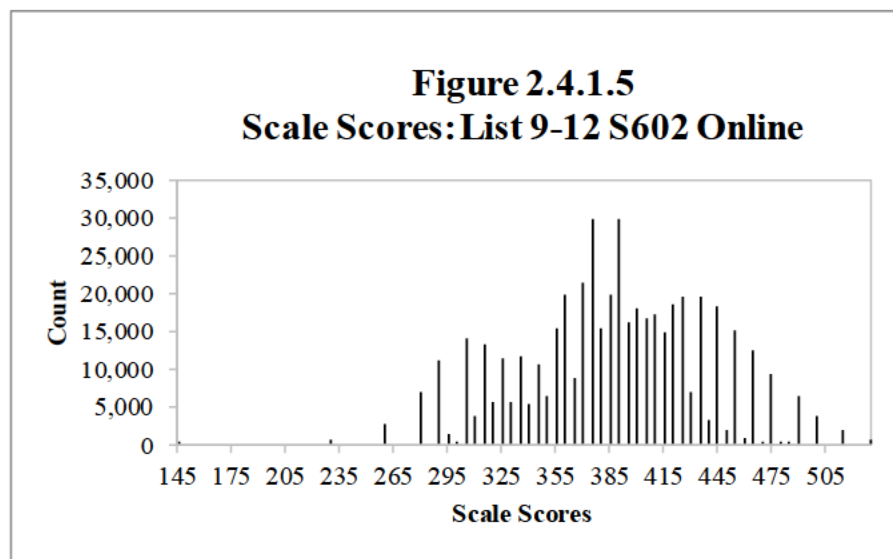
Table 2.4.1.5

Scale Score Descriptive Statistics: List 9–12 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	161,437	148	532	382.80	49.63
10	139,476	148	532	389.28	50.09
11	111,575	148	532	392.68	50.70
12	80,850	148	532	395.31	49.54
Total	493,338	148	532	388.92	50.21

Figure 2.4.1.5

Scale Scores: List 9–12 S602 Online



2.4.2 Reading

2.4.2.1 Grade 1

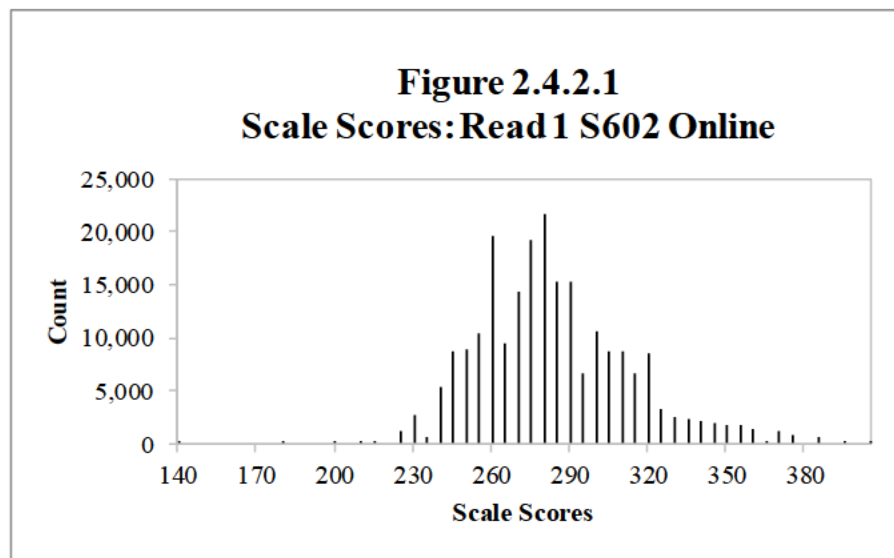
Table 2.4.2.1

Scale Score Descriptive Statistics: Read 1 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	223,101	141	406	285.94	29.24
Total	223,101	141	406	285.94	29.24

Figure 2.4.2.1

Scale Scores: Read 1 S602 Online



2.4.2.2 Grades 2–3

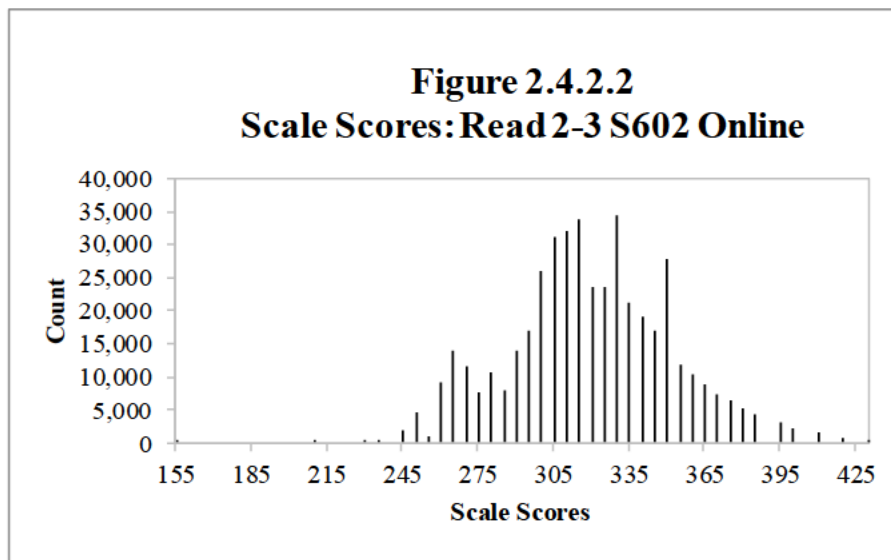
Table 2.4.2.2

Scale Score Descriptive Statistics: Read 2–3 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	230,737	158	431	316.60	27.73
3	221,419	158	431	326.63	35.08
Total	452,156	158	431	321.51	31.94

Figure 2.4.2.2

Scale Scores: Read 2–3 S602 Online



2.4.2.3 Grades 4–5

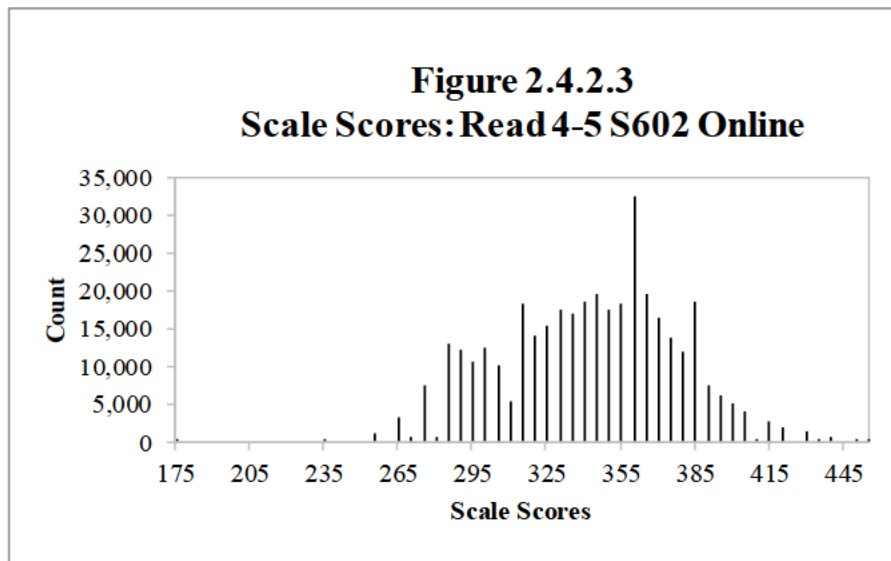
Table 2.4.2.3

Scale Score Descriptive Statistics: Read 4–5 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	204,270	175	459	343.29	33.83
5	169,851	175	459	346.34	35.97
Total	374,121	175	459	344.67	34.85

Figure 2.4.2.3

Scale Scores: Read 4–5 S602 Online



2.4.2.4 Grades 6–8

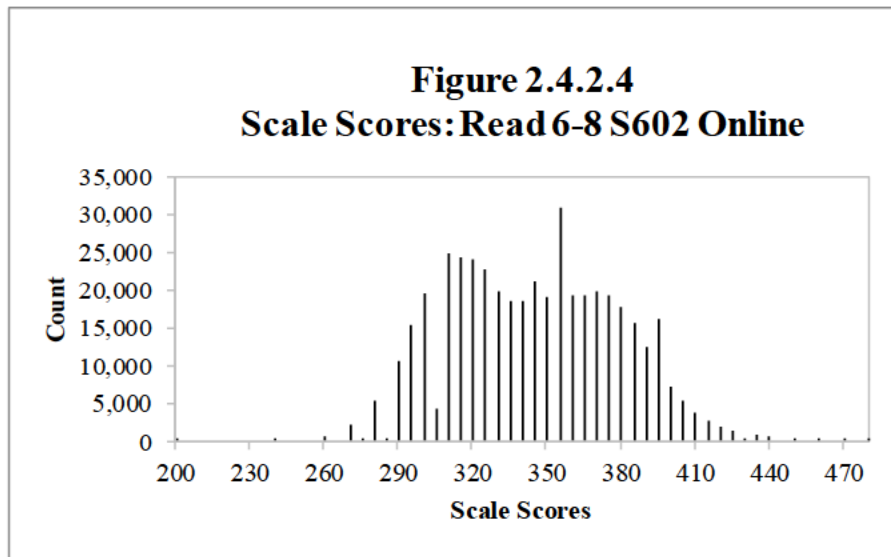
Table 2.4.2.4

Scale Score Descriptive Statistics: Read 6–8 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	145,493	200	482	339.95	30.90
7	151,390	200	482	347.41	33.60
8	149,602	200	482	354.25	35.80
Total	446,485	200	482	347.27	34.02

Figure 2.4.2.4

Scale Scores: Read 6–8 S602 Online



2.4.2.5 Grades 9–12

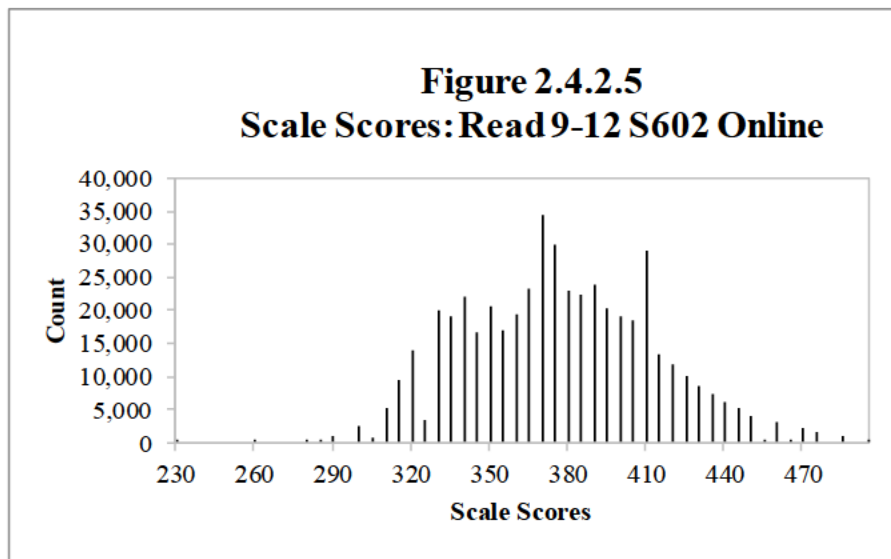
Table 2.4.2.5

Scale Score Descriptive Statistics: Read 9–12 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	160,871	262	497	373.39	33.96
10	138,099	233	497	379.01	35.25
11	110,332	233	497	381.78	37.27
12	79,923	262	497	383.08	37.43
Total	489,225	233	497	378.45	35.87

Figure 2.4.2.5

Scale Scores: Read 9–12 S602 Online



2.4.3 Writing

2.4.3.1 Grade 1

Table 2.4.3.1.1

Scale Score Descriptive Statistics: Writ 1 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	209,593	111	360	228.14	47.71
Total	209,593	111	360	228.14	47.71

Figure 2.4.3.1.1

Scale Scores: Writ 1 A S602 Online

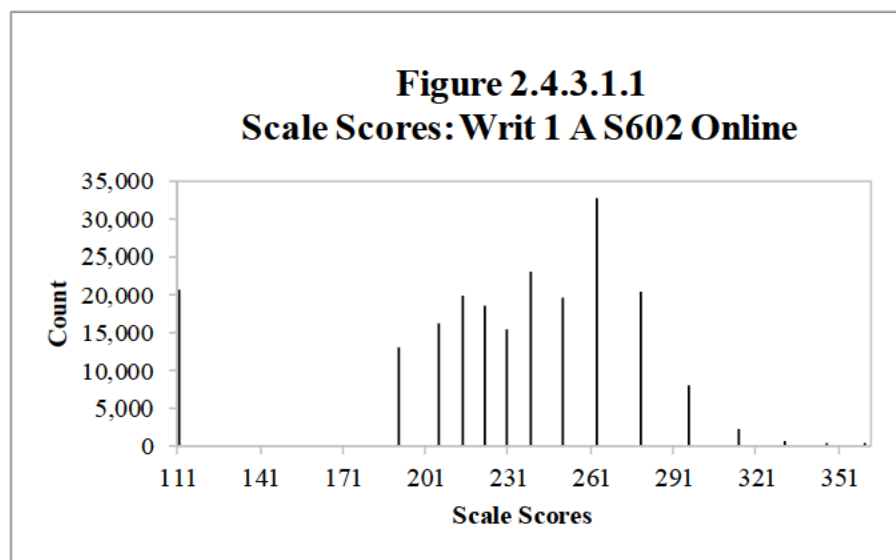


Table 2.4.3.1.2

Scale Score Descriptive Statistics: Writ 1 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	26,141	111	425	295.37	34.76
Total	26,141	111	425	295.37	34.76

Figure 2.4.3.1.2

Scale Scores: Writ 1 B/C S602 Online

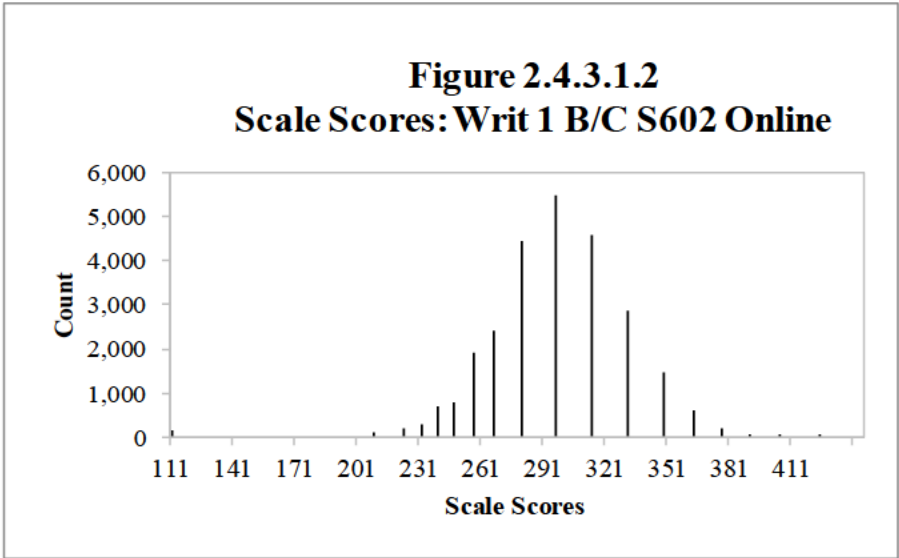


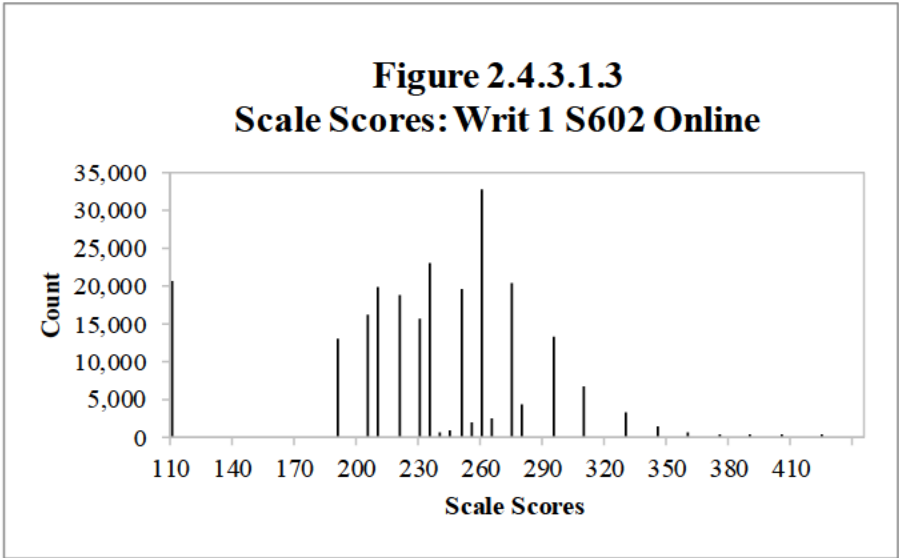
Table 2.4.3.1.3

Scale Score Descriptive Statistics: Writ 1 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	235,734	111	425	235.60	51.02
Total	235,734	111	425	235.60	51.02

Figure 2.4.3.1.3

Scale Scores: Writ 1 S602 Online



2.4.3.2 Grades 2–3

Table 2.4.3.2.1

Scale Score Descriptive Statistics: Writ 2–3 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	87,830	133	387	234.32	54.64
3	73,330	133	401	243.55	56.25
Total	161,160	133	401	238.52	55.57

Figure 2.4.3.2.1

Scale Scores: Writ 2–3 A S602 Online

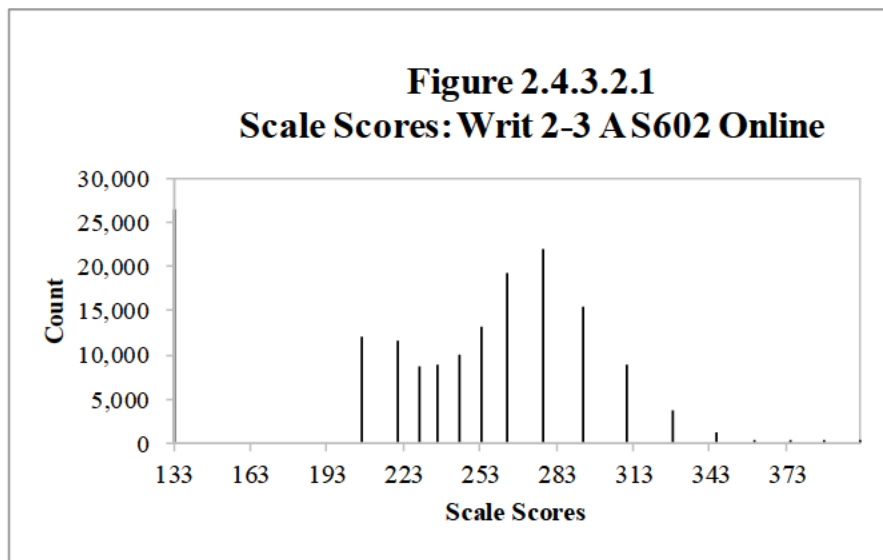


Table 2.4.3.2.2

Scale Score Descriptive Statistics: Writ 2–3 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	159,427	133	459	297.27	39.64
3	164,235	133	459	318.57	32.62
Total	323,662	133	459	308.08	37.78

Figure 2.4.3.2.2

Scale Scores: Writ 2–3 B/C S602 Online

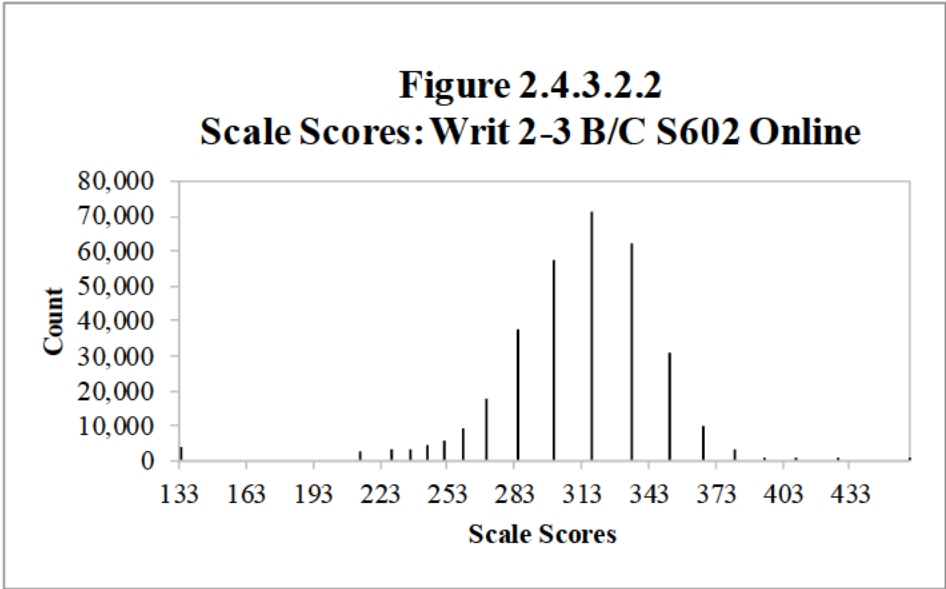


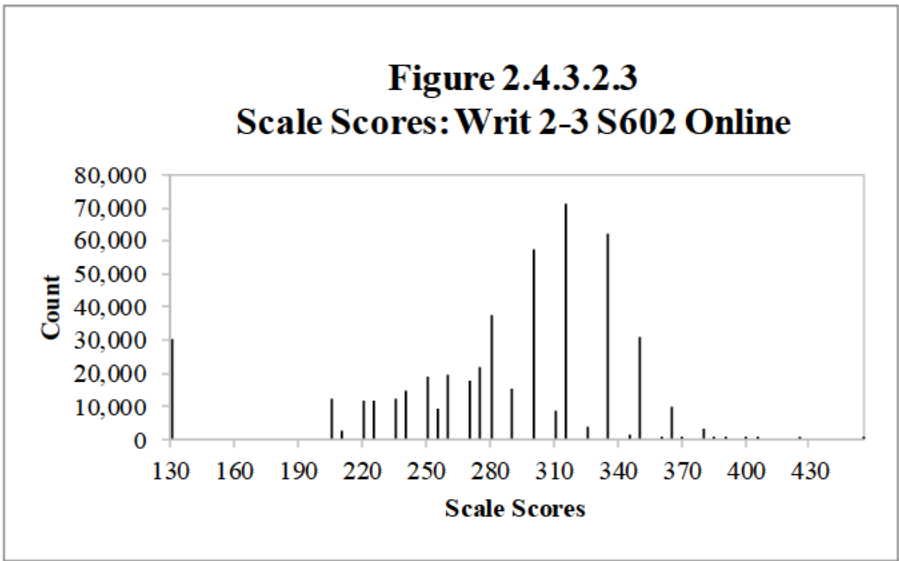
Table 2.4.3.2.3

Scale Score Descriptive Statistics: Writ 2–3 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	247,257	133	459	274.91	54.60
3	237,565	133	459	295.42	53.97
Total	484,822	133	459	284.96	55.25

Figure 2.4.3.2.3

Scale Scores: Writ 2–3 S602 Online



2.4.3.3 Grades 4–5

Table 2.4.3.3.1

Scale Score Descriptive Statistics: Writ 4–5 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	56,044	155	417	252.48	55.95
5	53,170	155	404	262.62	55.87
Total	109,214	155	417	257.42	56.14

Figure 2.4.3.3.1

Scale Scores: Writ 4–5 A S602 Online

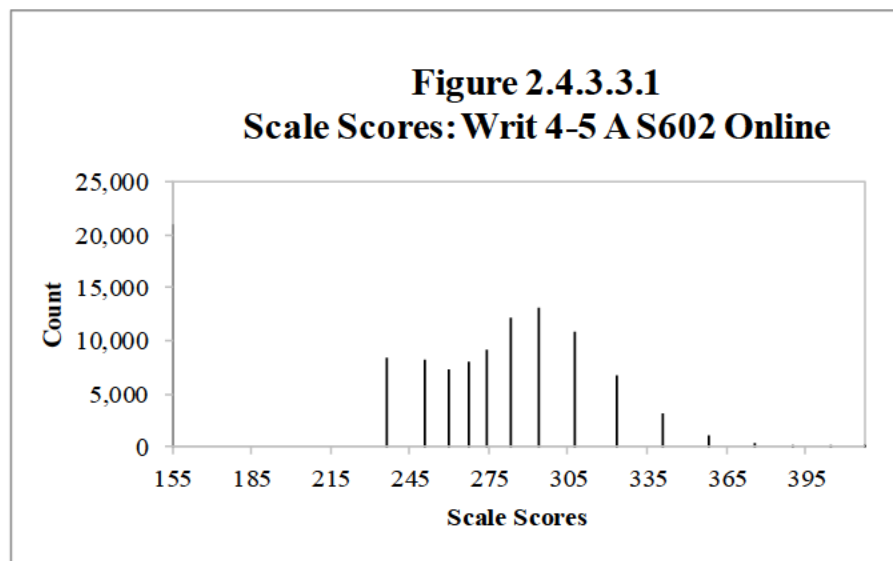


Table 2.4.3.3.2

Scale Score Descriptive Statistics: Writ 4–5 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	151,995	155	475	340.42	36.15
5	120,478	155	475	352.23	32.25
Total	272,473	155	475	345.65	34.97

Figure 2.4.3.3.2

Scale Scores: Writ 4–5 B/C S602 Online

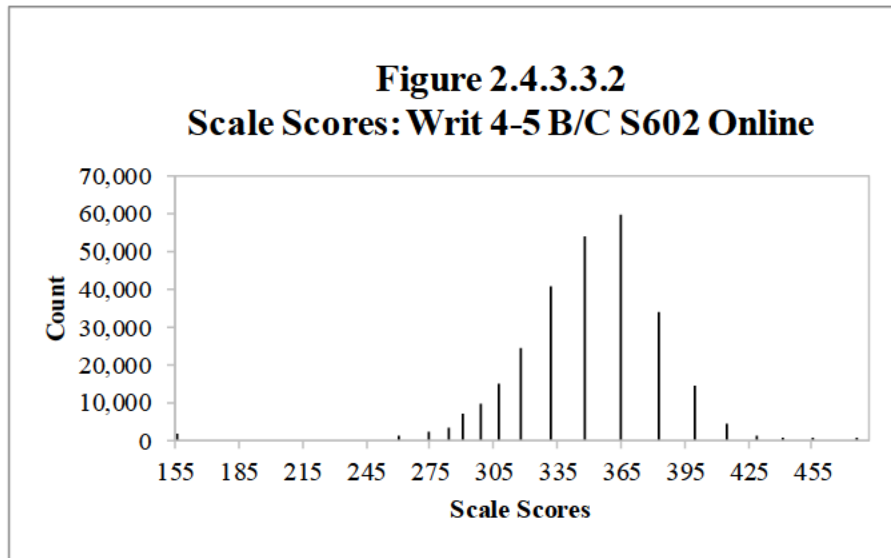


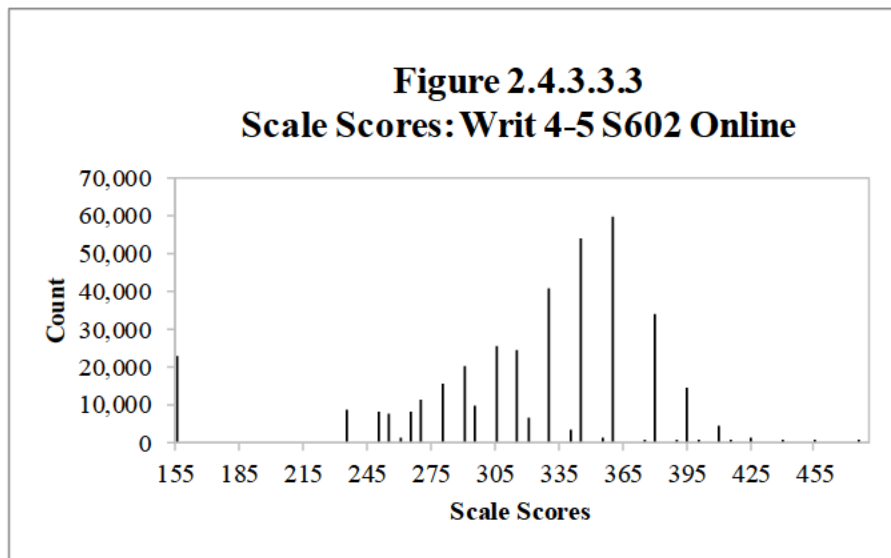
Table 2.4.3.3.3

Scale Score Descriptive Statistics: Writ 4–5 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	208,039	155	475	316.73	57.62
5	173,648	155	475	324.79	58.17
Total	381,687	155	475	320.40	58.01

Figure 2.4.3.3.3

Scale Scores: Writ 4–5 S602 Online



2.4.3.4 Grades 6–8

Table 2.4.3.4.1

Scale Score Descriptive Statistics: Writ 6–8 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	61,057	188	409	273.00	38.93
7	72,746	188	423	283.50	39.92
8	73,194	188	443	289.60	41.03
Total	206,997	188	443	282.56	40.58

Figure 2.4.3.4.1

Scale Scores: Writ 6–8 A S602 Online

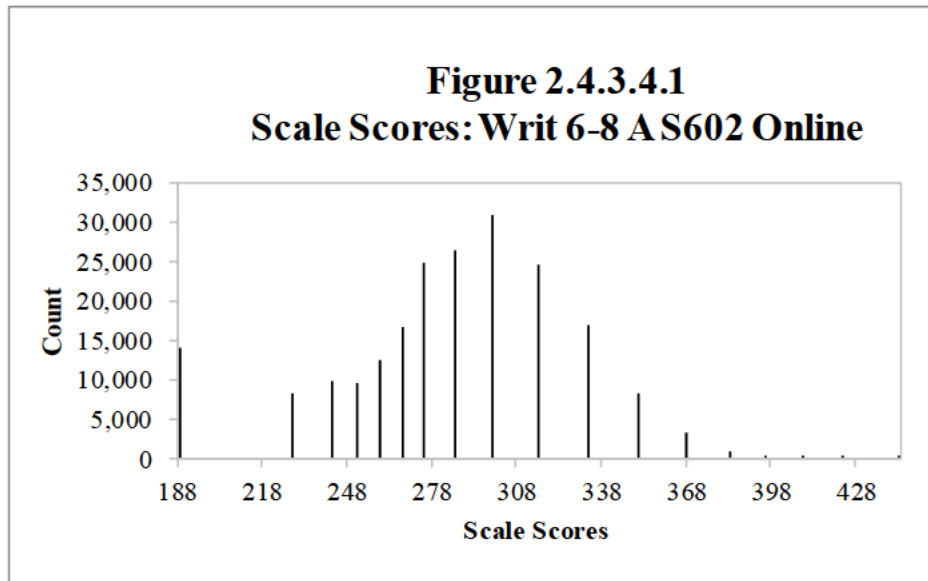


Table 2.4.3.4.2

Scale Score Descriptive Statistics: Writ 6–8 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	91,844	188	440	327.88	29.46
7	86,756	188	440	341.94	28.73
8	83,536	188	460	351.92	28.76
Total	262,136	188	460	340.19	30.64

Figure 2.4.3.4.2

Scale Scores: Writ 6–8 B/C S602 Online

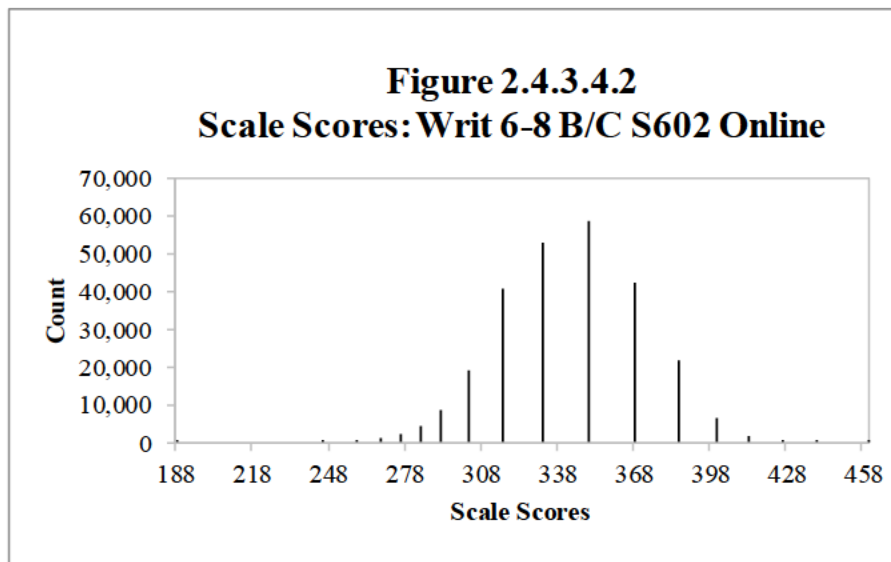


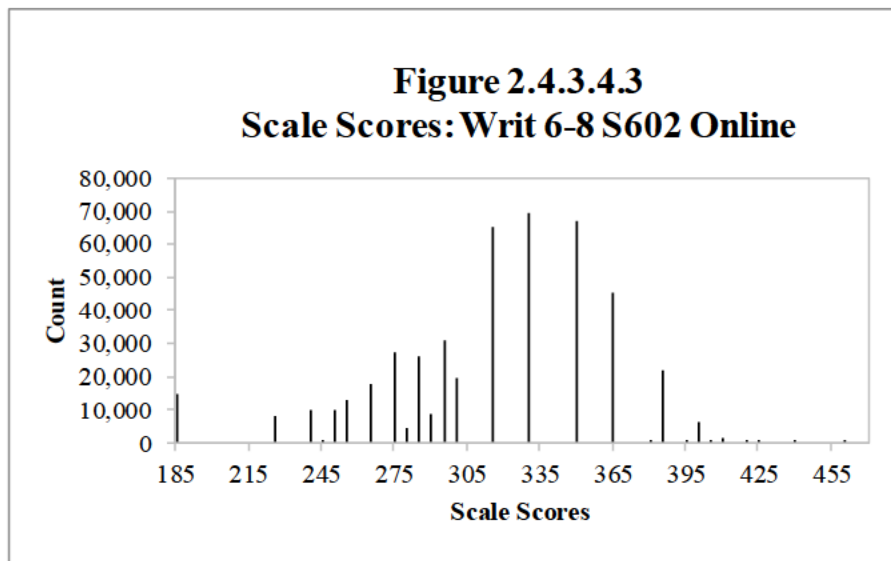
Table 2.4.3.4.3

Scale Score Descriptive Statistics: Writ 6–8 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	152,901	188	440	305.96	43.00
7	159,502	188	440	315.29	44.98
8	156,730	188	460	322.81	46.84
Total	469,133	188	460	314.76	45.50

Figure 2.4.3.4.3

Scale Scores: Writ 6–8 S602 Online



2.4.3.5 Grades 9–12

Table 2.4.3.5.1

Scale Score Descriptive Statistics: Writ 9–12 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	70,048	232	474	308.33	43.51
10	52,662	232	454	315.57	40.62
11	41,969	232	454	322.33	39.87
12	27,325	232	474	325.02	39.64
Total	192,004	232	474	315.75	41.91

Figure 2.4.3.5.1

Scale Scores: Writ 9–12 A S602 Online

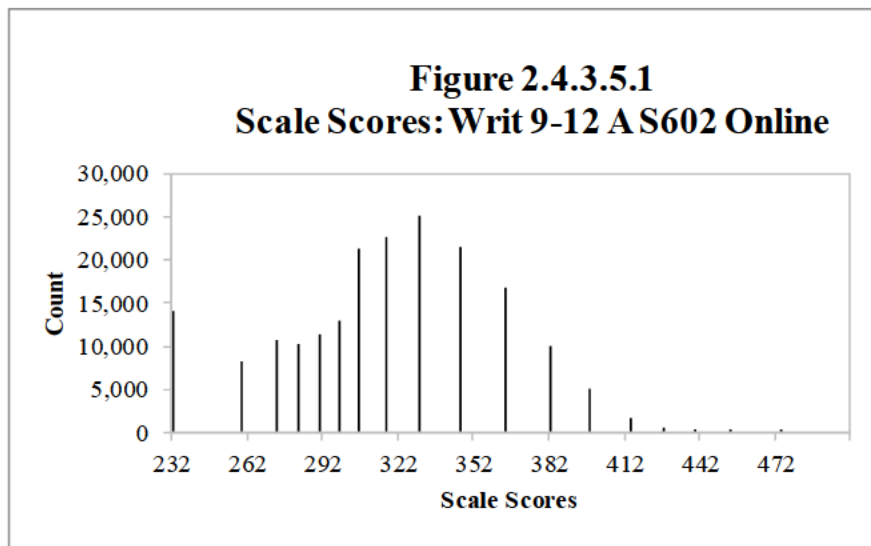


Table 2.4.3.5.2

Scale Score Descriptive Statistics: Writ 9–12 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	98,574	232	475	354.04	33.95
10	91,888	232	506	356.87	33.56
11	73,351	232	475	360.52	32.82
12	56,123	232	475	358.73	33.27
Total	319,936	232	506	357.16	33.55

Figure 2.4.3.5.2

Scale Scores: Writ 9–12 B/C S602 Online

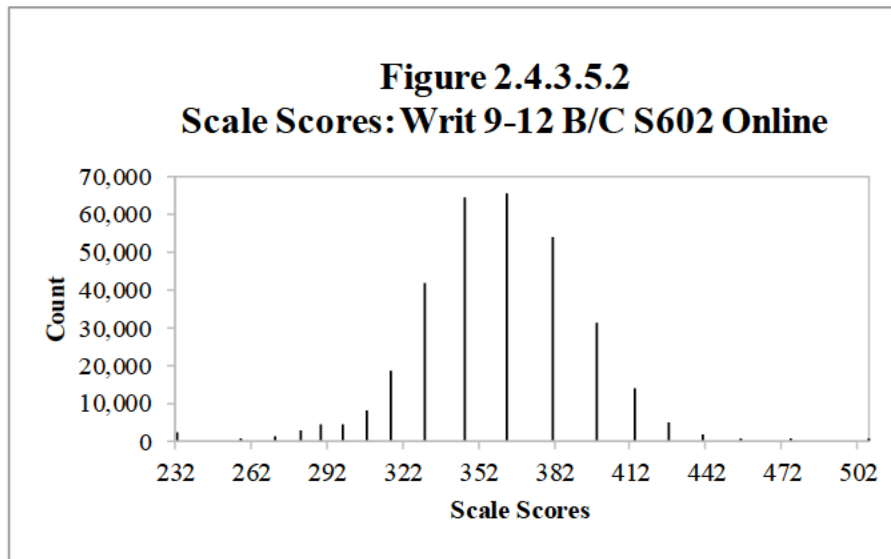


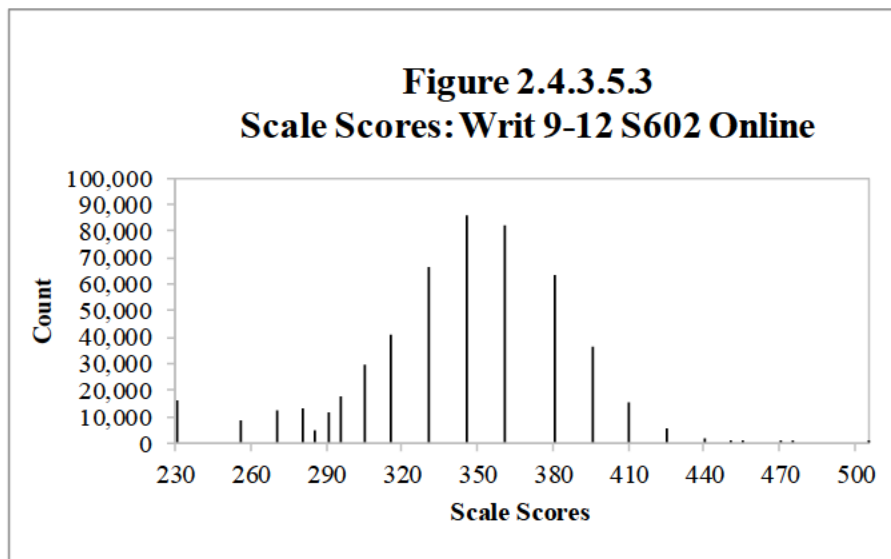
Table 2.4.3.5.3

Scale Score Descriptive Statistics: Writ 9–12 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	168,622	232	475	335.05	44.36
10	144,550	232	506	341.82	41.38
11	115,320	232	475	346.62	40.01
12	83,448	232	475	347.69	38.85
Total	511,940	232	506	341.63	42.00

Figure 2.4.3.5.3

Scale Scores: Writ 9–12 S602 Online



2.4.4 Speaking

2.4.4.1 Grade 1

Table 2.4.4.1.1

Scale Score Descriptive Statistics: Spek 1 Pre-A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	14,805	106	167	143.11	25.18
Total	14,805	106	167	143.11	25.18

Figure 2.4.4.1.1

Scale Scores: Spek 1 Pre-A S602 Online

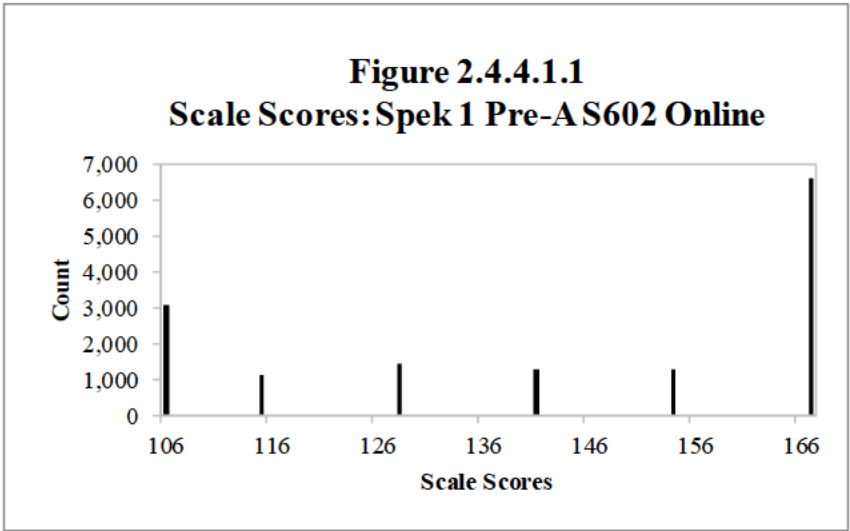


Table 2.4.4.1.2

Scale Score Descriptive Statistics: Spek 1 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	108,591	106	373	217.64	57.54
Total	108,591	106	373	217.64	57.54

Figure 2.4.4.1.2

Scale Scores: Spek 1 A S602 Online

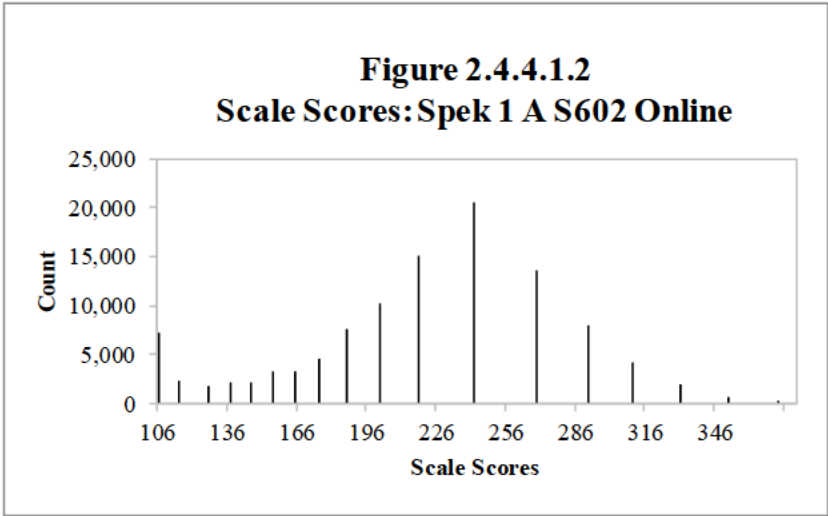


Table 2.4.4.1.3

Scale Score Descriptive Statistics: Spek 1 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	91,009	106	403	266.72	46.69
Total	91,009	106	403	266.72	46.69

Figure 2.4.4.1.3

Scale Scores: Spek 1 B/C S602 Online

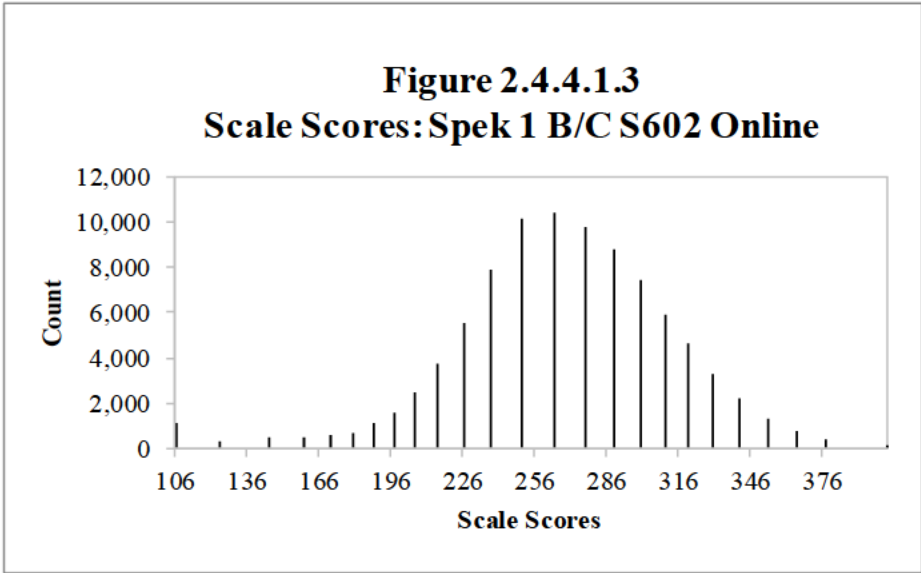


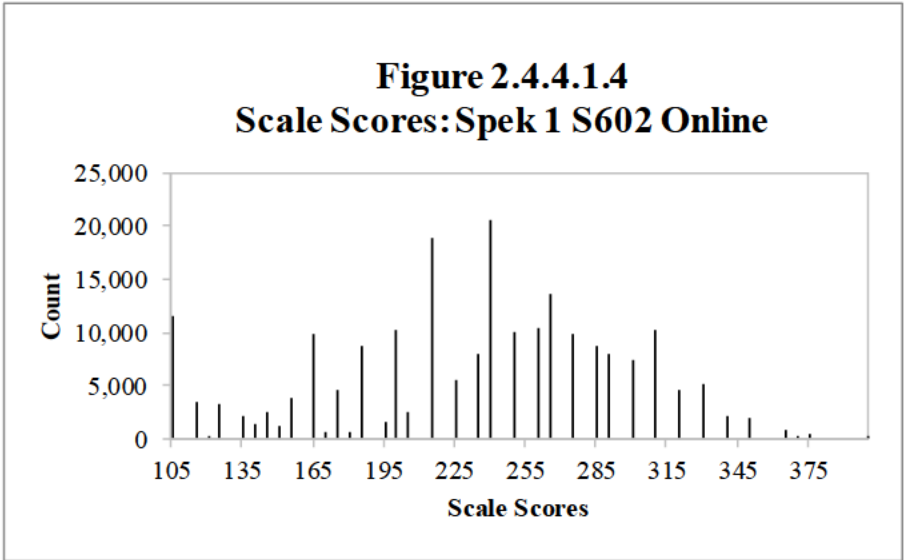
Table 2.4.4.1.4

Scale Score Descriptive Statistics: Spek 1 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	214,405	106	403	233.33	61.69
Total	214,405	106	403	233.33	61.69

Figure 2.4.4.1.4

Scale Scores: Spek 1 S602 Online



2.4.4.2 Grades 2–3

Table 2.4.4.2.1

Scale Score Descriptive Statistics: Spek 2–3 Pre-A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	9,393	118	164	146.60	20.01
3	20,574	118	164	147.68	19.70
Total	29,967	118	164	147.34	19.80

Figure 2.4.4.2.1

Scale Scores: Spek 2–3 Pre-A S602 Online

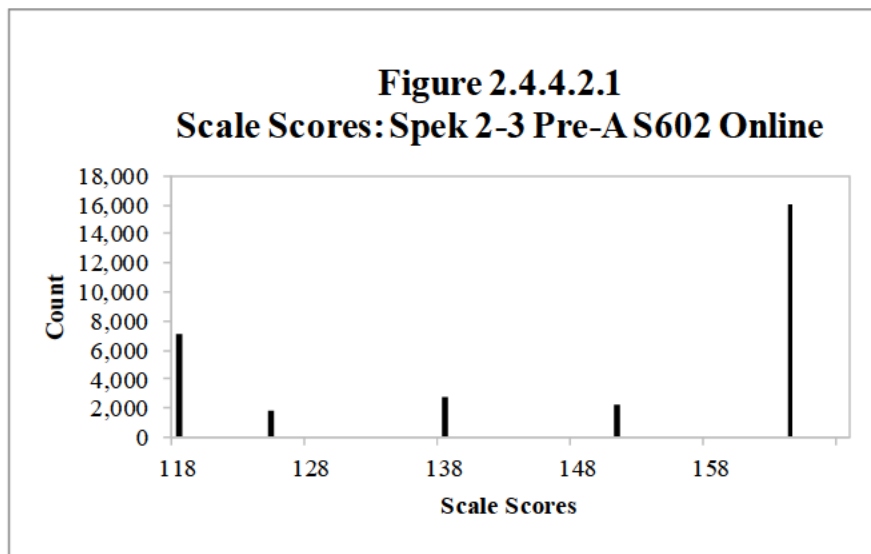


Table 2.4.4.2.2

Scale Score Descriptive Statistics: Spek 2–3 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	74,659	118	391	225.79	58.40
3	67,665	118	391	250.98	57.39
Total	142,324	118	391	237.76	59.27

Figure 2.4.4.2.2

Scale Scores: Spek 2–3 A S602 Online

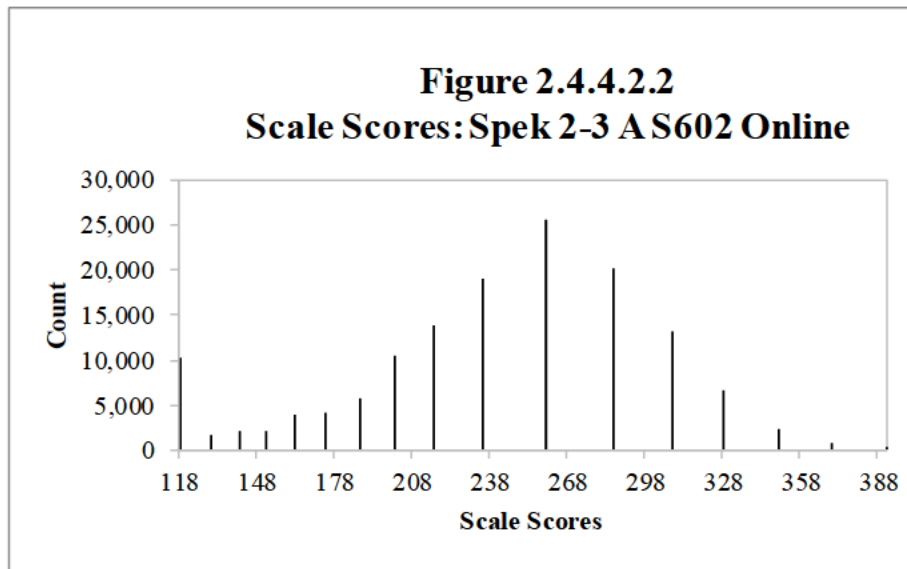


Table 2.4.4.2.3

Scale Score Descriptive Statistics: Spek 2–3 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	141,555	118	425	281.40	46.20
3	131,745	118	425	302.88	43.21
Total	273,300	118	425	291.75	46.05

Figure 2.4.4.2.3

Scale Scores: Spek 2–3 B/C S602 Online

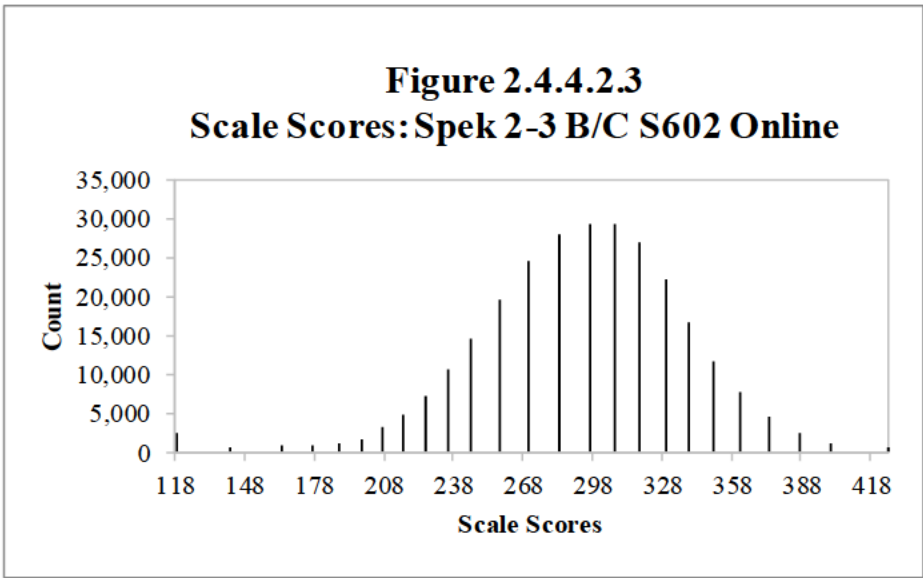


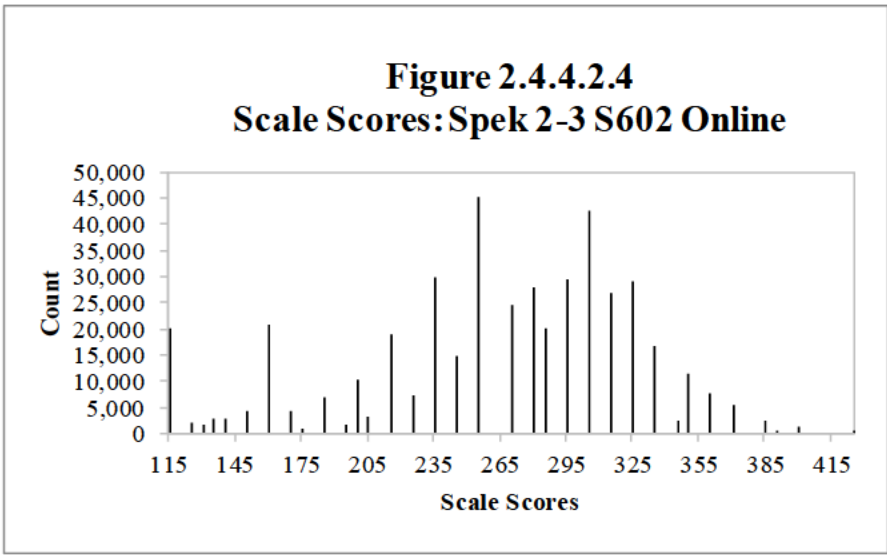
Table 2.4.4.2.4

Scale Score Descriptive Statistics: Spek 2–3 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	225,607	118	425	257.38	60.73
3	219,984	118	425	272.40	65.72
Total	445,591	118	425	264.80	63.69

Figure 2.4.4.2.4

Scale Scores: Spek 2–3 S602 Online



2.4.4.3 Grades 4–5

Table 2.4.4.3.1

Scale Score Descriptive Statistics: Spek 4–5 Pre-A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	4,046	130	190	163.78	24.63
5	8,610	130	190	166.71	24.20
Total	12,656	130	190	165.78	24.37

Figure 2.4.4.3.1

Scale Scores: Spek 4–5 Pre-A S602 Online

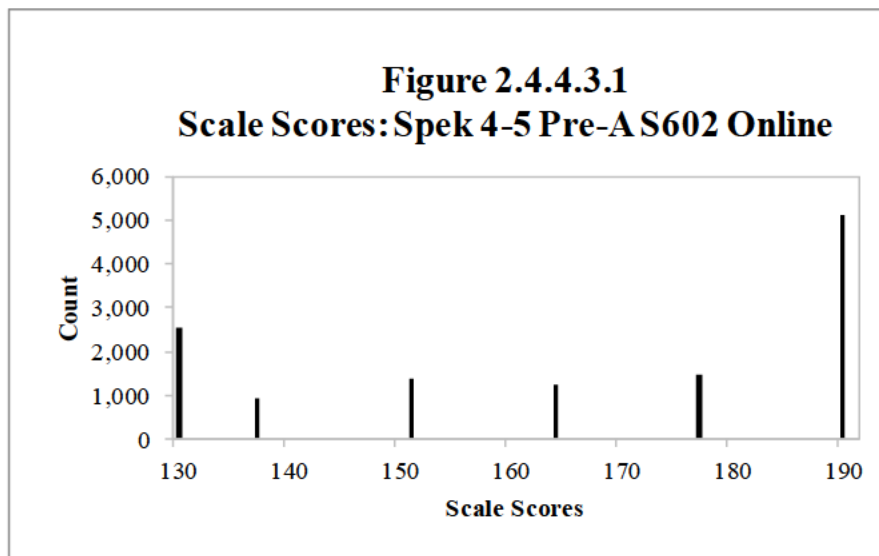


Table 2.4.4.3.2

Scale Score Descriptive Statistics: Spek 4–5 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	43,530	130	431	244.86	57.08
5	32,772	130	431	249.03	57.04
Total	76,302	130	431	246.65	57.10

Figure 2.4.4.3.2

Scale Scores: Spek 4–5 A S602 Online

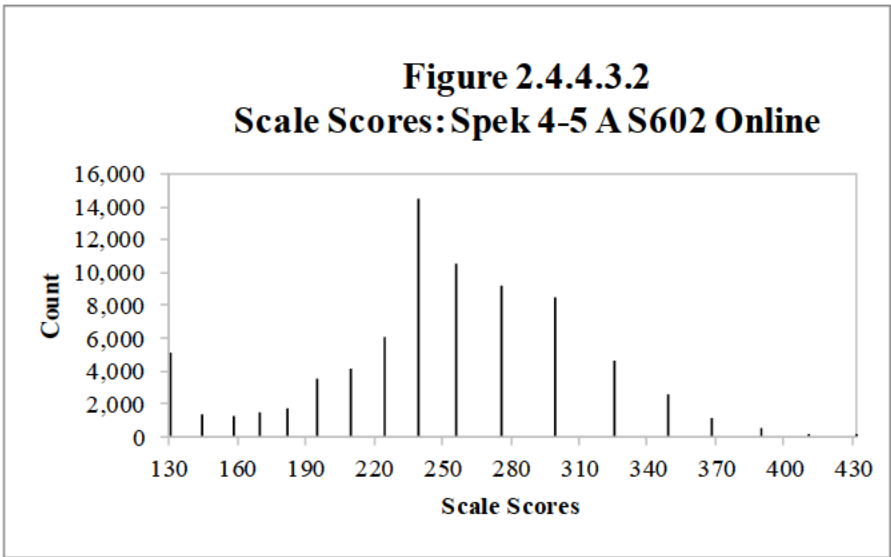


Table 2.4.4.3.3

Scale Score Descriptive Statistics: Spek 4–5 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	158,569	130	459	329.21	48.54
5	130,202	130	459	330.54	48.88
Total	288,771	130	459	329.81	48.70

Figure 2.4.4.3.3

Scale Scores: Spek 4–5 B/C S602 Online

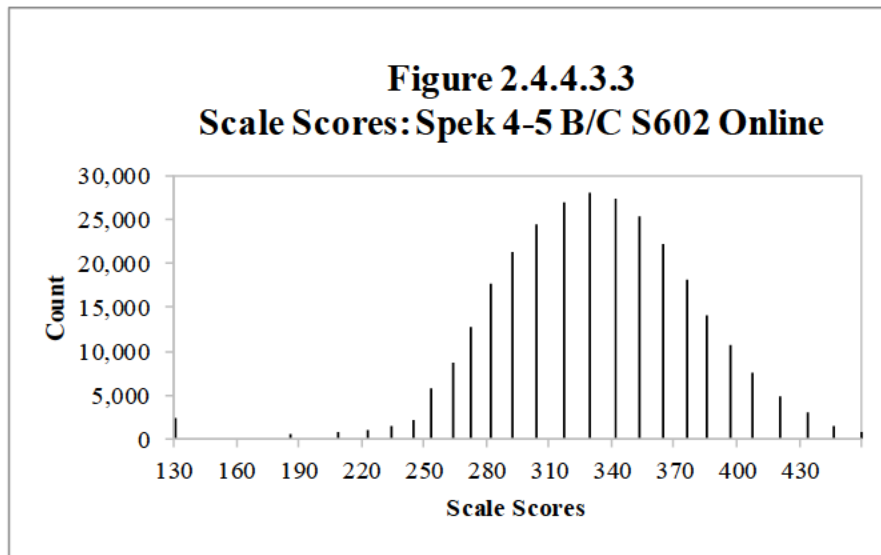


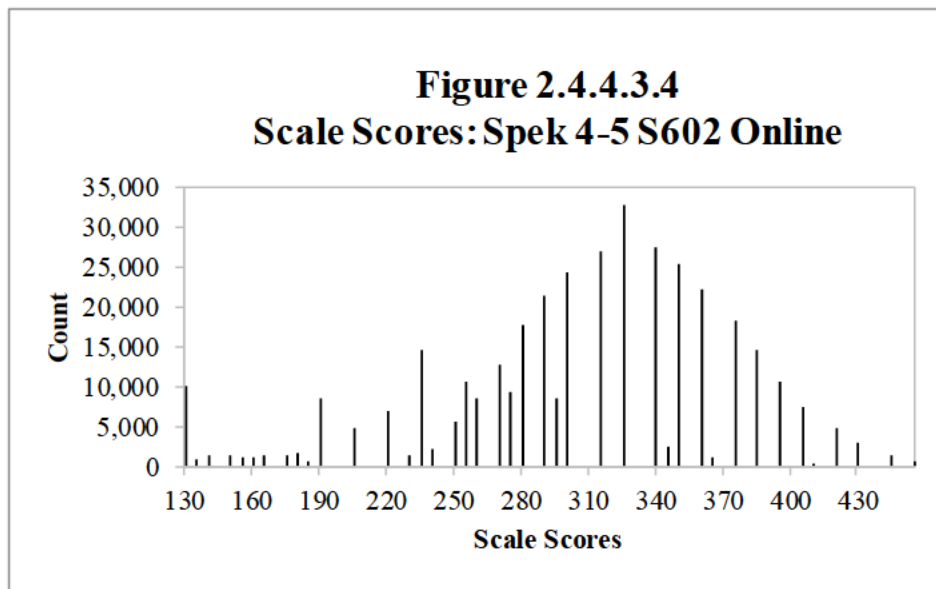
Table 2.4.4.3.4

Scale Score Descriptive Statistics: Spek 4–5 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	206,145	130	459	308.16	64.09
5	171,584	130	459	306.75	67.18
Total	377,729	130	459	307.52	65.52

Figure 2.4.4.3.4

Scale Scores: Spek 4–5 S602 Online



2.4.4.4 Grades 6–8

Table 2.4.4.4.1

Scale Score Descriptive Statistics: Spek 6–8 Pre-A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	6,152	148	212	192.51	24.87
7	10,204	148	212	193.82	24.39
8	14,536	148	212	195.43	23.81
Total	30,892	148	212	194.32	24.24

Figure 2.4.4.4.1

Scale Scores: Spek 6–8 Pre-A S602 Online

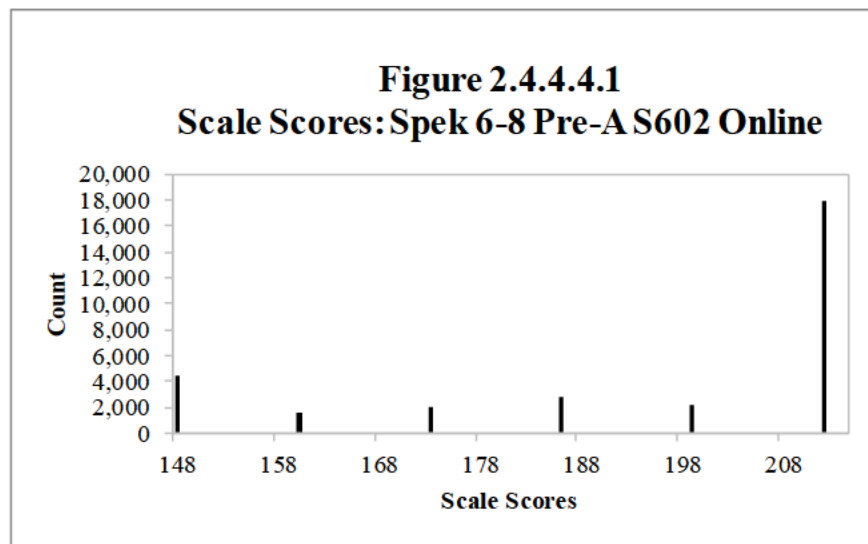


Table 2.4.4.4.2

Scale Score Descriptive Statistics: Spek 6–8 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	35,089	148	432	254.19	54.00
7	25,694	148	432	251.87	53.17
8	47,042	148	432	278.42	57.70
Total	107,825	148	432	264.21	56.85

Figure 2.4.4.4.2

Scale Scores: Spek 6–8 A S602 Online

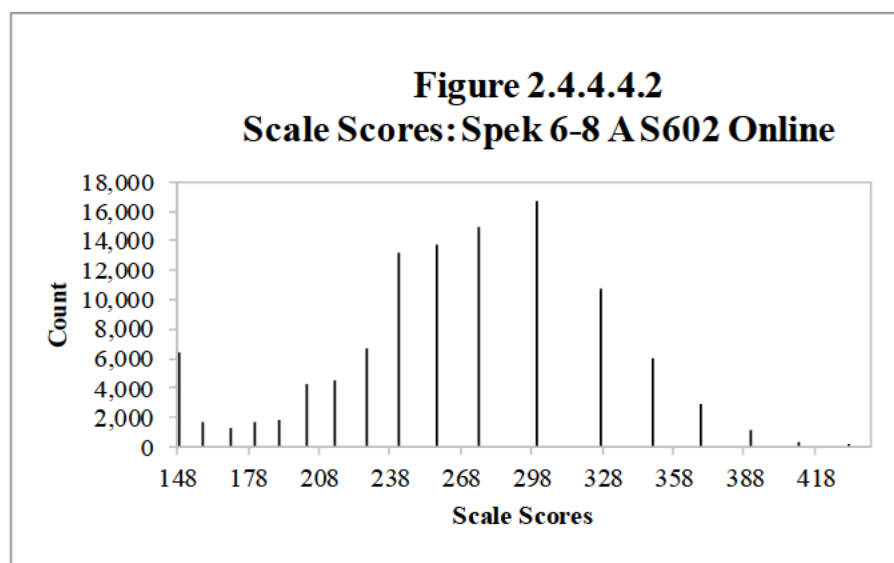


Table 2.4.4.4.3

Scale Score Descriptive Statistics: Spek 6–8 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	104,526	148	463	321.34	48.01
7	115,424	148	463	324.98	51.83
8	89,410	148	463	340.56	49.56
Total	309,360	148	463	328.25	50.55

Figure 2.4.4.4.3

Scale Scores: Spek 6–8 B/C S602 Online

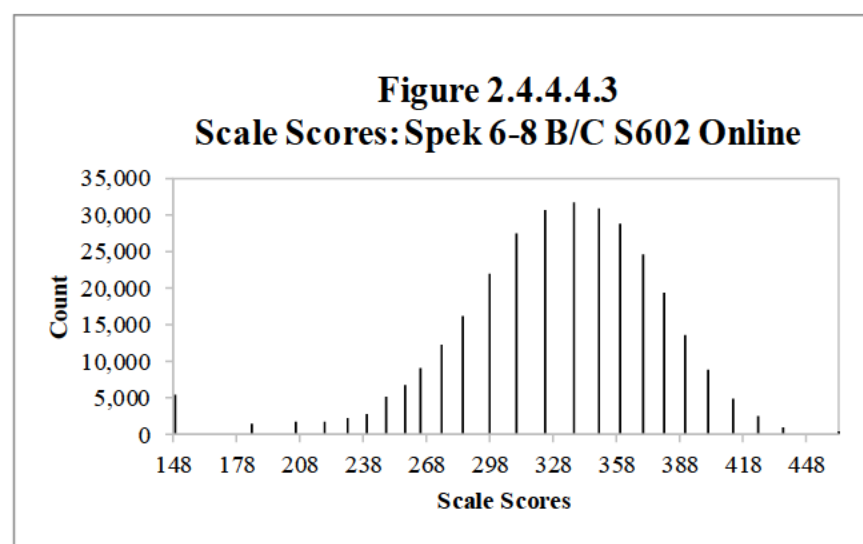


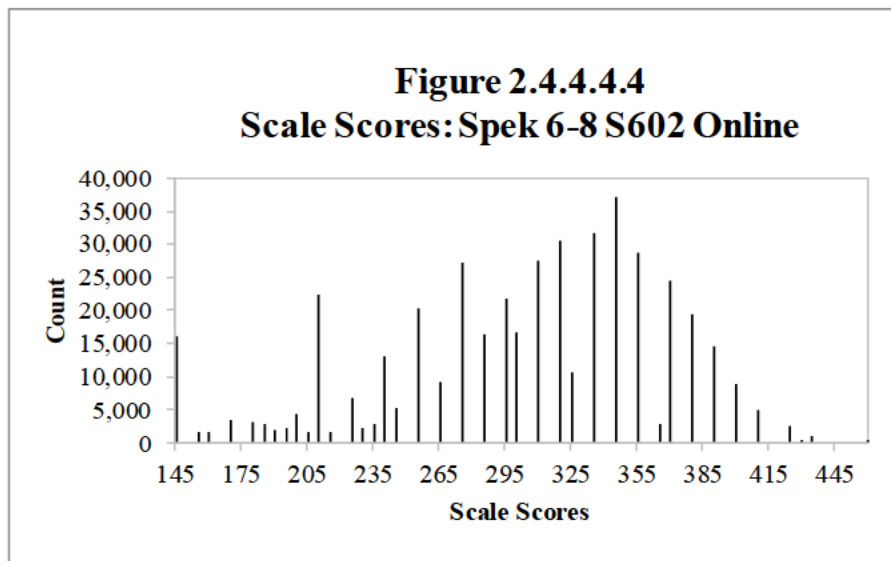
Table 2.4.4.4.4

Scale Score Descriptive Statistics: Spek 6–8 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	145,767	148	463	299.74	60.83
7	151,322	148	463	303.72	64.69
8	150,988	148	463	307.23	68.31
Total	448,077	148	463	303.61	64.80

Figure 2.4.4.4.4

Scale Scores: Spek 6–8 S602 Online



2.4.4.5 Grades 9–12

Table 2.4.4.5.1

Scale Score Descriptive Statistics: Spek 9–12 Pre-A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	9,161	172	219	203.43	18.97
10	9,342	172	219	207.95	17.22
11	11,193	172	219	209.42	16.73
12	8,498	172	219	210.27	16.52
Total	38,194	172	219	207.81	17.56

Figure 2.4.4.5.1

Scale Scores: Spek 9–12 Pre-A S602 Online

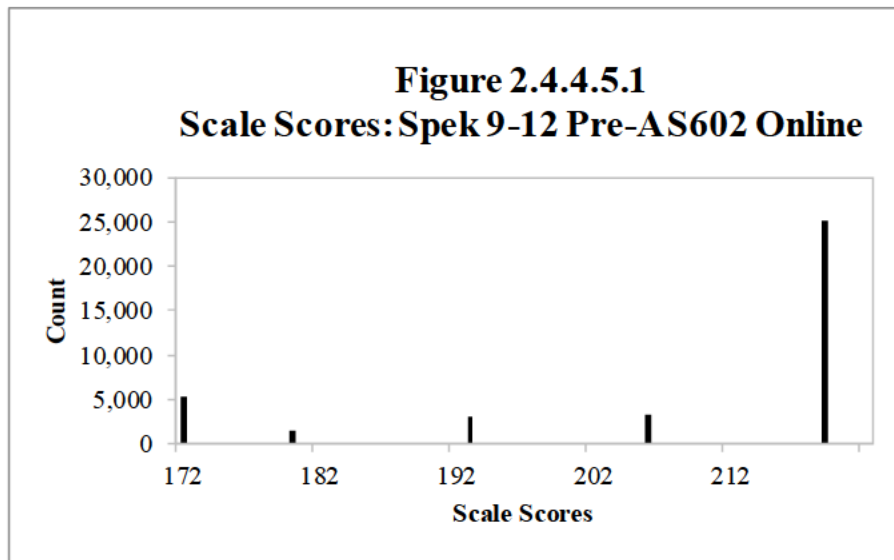


Table 2.4.4.5.2

Scale Score Descriptive Statistics: Spek 9–12 A S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	94,373	172	444	281.00	56.47
10	60,308	172	444	283.35	55.02
11	24,935	172	444	280.47	52.84
12	39,411	172	444	304.30	56.29
Total	219,027	172	444	285.78	56.32

Figure 2.4.4.5.2

Scale Scores: Spek 9–12 A S602 Online

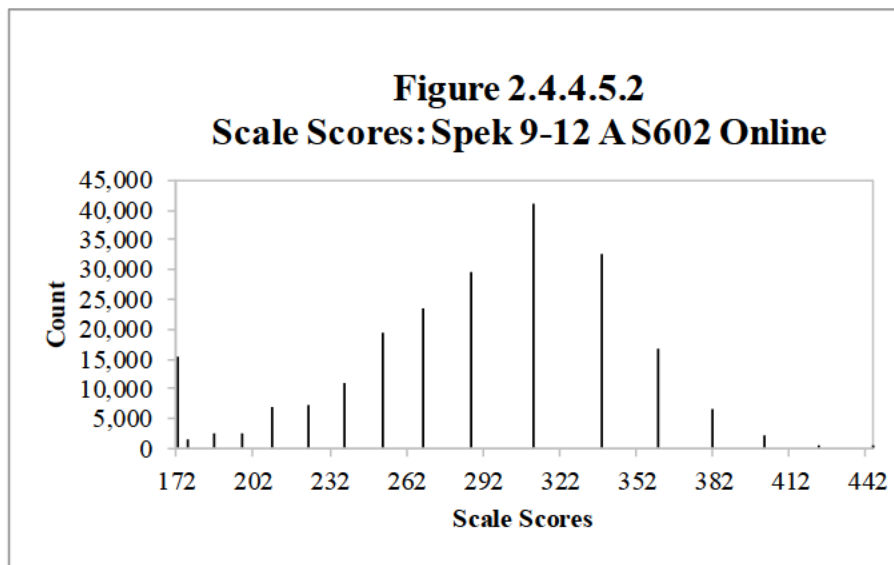


Table 2.4.4.5.3

Scale Score Descriptive Statistics: Spek 9–12 B/C S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	61,928	172	476	334.11	46.66
10	71,692	172	476	334.83	48.93
11	75,932	172	476	330.95	52.38
12	34,390	172	476	342.79	49.14
Total	243,942	172	476	334.56	49.64

Figure 2.4.4.5.3

Scale Scores: Spek 9–12 B/C S602 Online

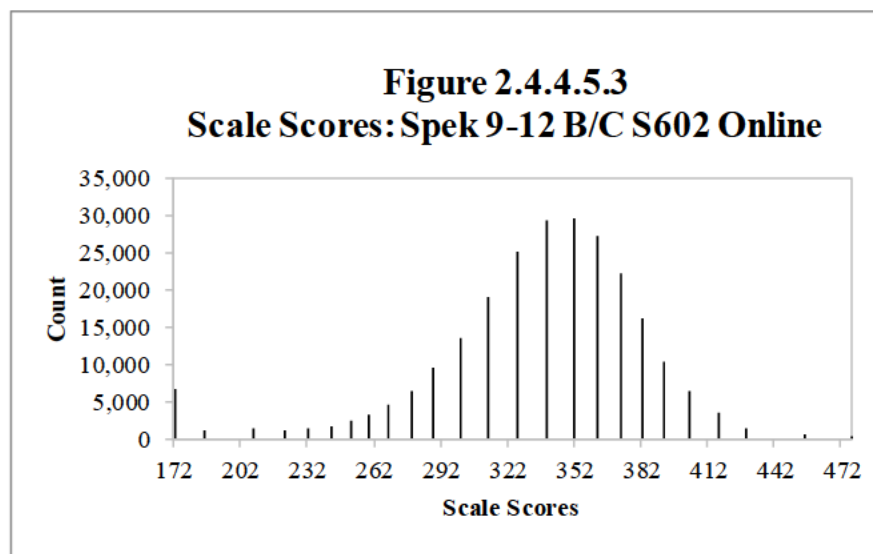


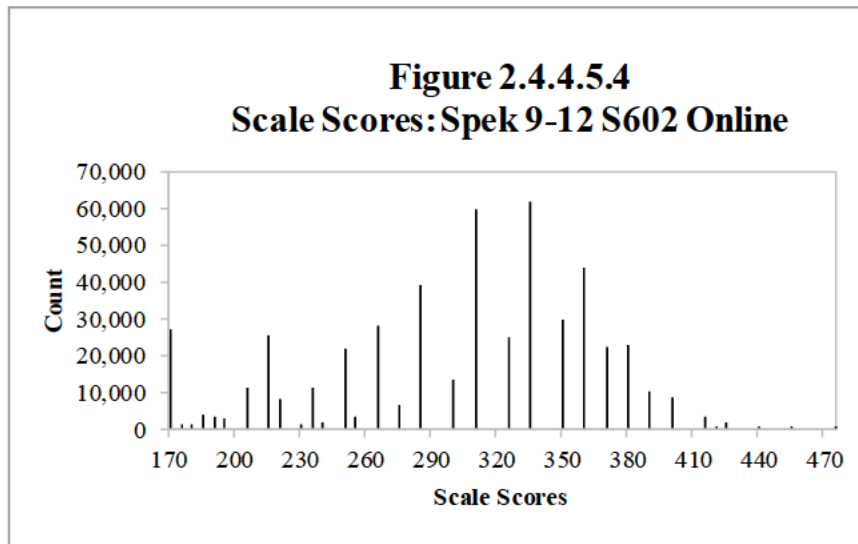
Table 2.4.4.5.4

Scale Score Descriptive Statistics: Spek 9–12 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	165,462	172	476	296.58	61.64
10	141,342	172	476	304.48	61.64
11	112,060	172	476	307.58	63.28
12	82,299	172	476	310.68	63.61
Total	501,163	172	476	303.58	62.56

Figure 2.4.4.5.4

Scale Scores: Spek 9–12 S602 Online



2.5 Proficiency Level Distributions

The figures and tables in this section provide information about the proficiency level distributions of the students who took each test form based on their performance by grade-level cluster. For Writing and Speaking, we also present that information by grade-level cluster and tier.

In the tables presented in this section, each row shows the following information, by grade (G#) and by total for the grade-level cluster:

- The WIDA proficiency level designation (1–6)
- The number of students (count) whose performance on the test form placed them into that proficiency level in the tested domain
- The percentage of students, out of the total number of students taking the form, who were placed into that proficiency level in the tested domain

In the figures, the horizontal axis shows the six WIDA proficiency levels. The vertical axis shows the percentage of students. Each bar shows the percentage of students who were placed into each proficiency level in the domain on this test form.

Note that WIDA intends for students who are just beginning to learn English to take the Speaking Pre-A tier; therefore, WIDA does not expect students assigned to this tier to show proficiency above PL 1.

2.5.1 Listening

2.5.1.1 Grade 1

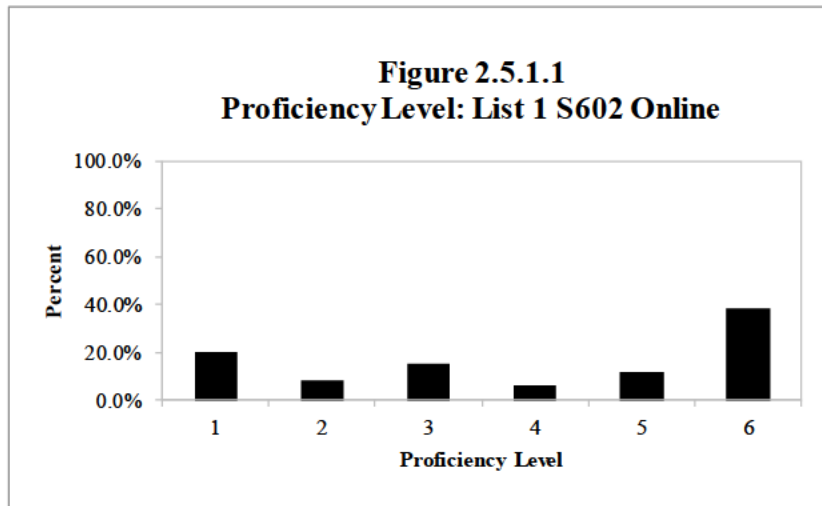
Table 2.5.1.1

Proficiency Level Distribution: List 1 S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	43,121	20.19%	43,121	20.19%
2	17,086	8.00%	17,086	8.00%
3	32,258	15.11%	32,258	15.11%
4	13,443	6.29%	13,443	6.29%
5	26,003	12.18%	26,003	12.18%
6	81,644	38.23%	81,644	38.23%
Total	213,555	100.00%	213,555	100.00%

Figure 2.5.1.1

Proficiency Level: List 1 S602 Online



2.5.1.2 Grades 2–3

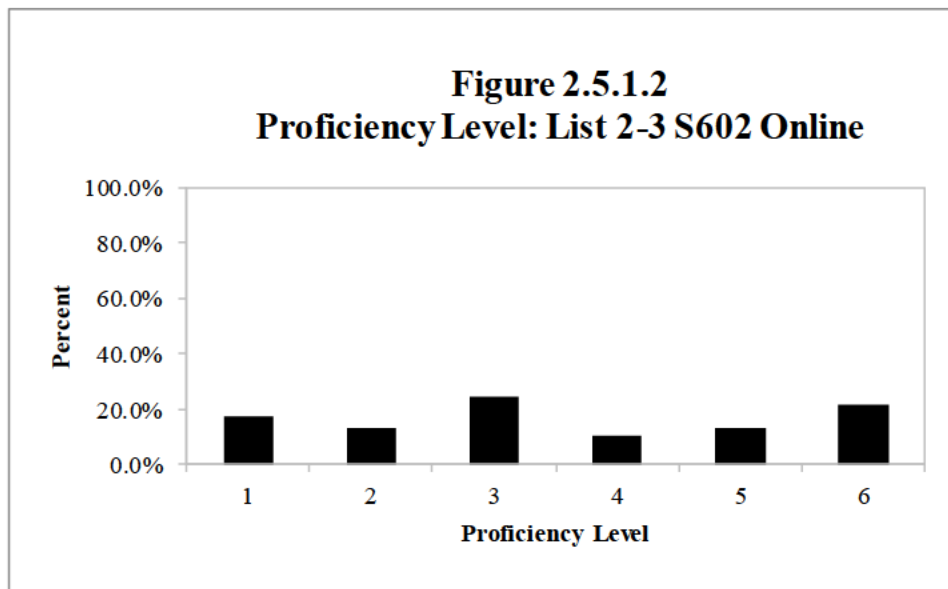
Table 2.5.1.2

Proficiency Level Distribution: List 2–3 S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	37,757	17.07%	38,000	17.61%	75,757	17.34%
2	30,380	13.74%	26,594	12.33%	56,974	13.04%
3	56,073	25.35%	50,531	23.42%	106,604	24.40%
4	23,789	10.76%	21,105	9.78%	44,894	10.27%
5	29,093	13.15%	28,908	13.40%	58,001	13.27%
6	44,090	19.93%	50,608	23.46%	94,698	21.67%
Total	221,182	100.00%	215,746	100.00%	436,928	100.00%

Figure 2.5.1.2

Proficiency Level: List 2–3 S602 Online



2.5.1.3 Grades 4–5

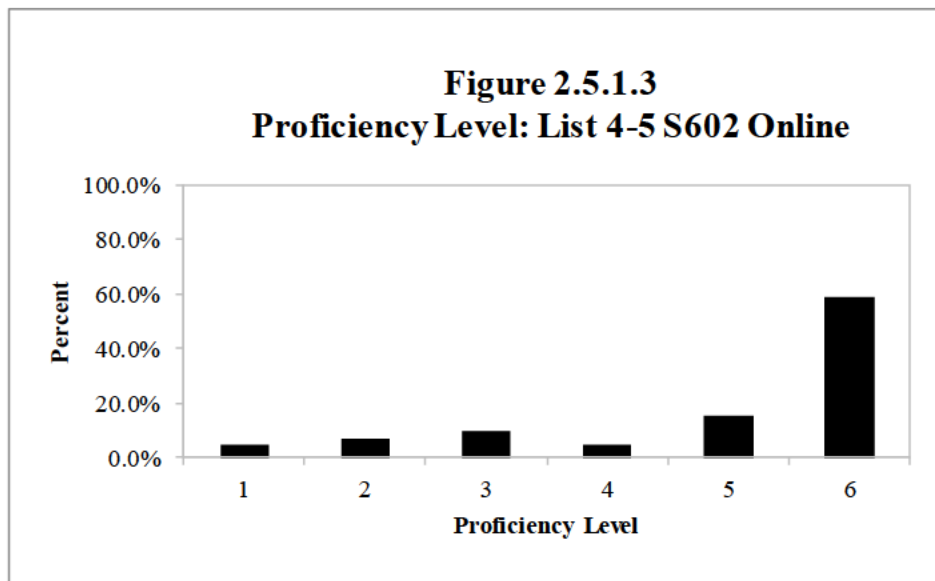
Table 2.5.1.3

Proficiency Level Distribution: List 4–5 S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	6,508	3.20%	10,323	6.08%	16,831	4.51%
2	12,751	6.26%	13,297	7.84%	26,048	6.98%
3	23,510	11.55%	13,961	8.23%	37,471	10.04%
4	8,954	4.40%	8,372	4.93%	17,326	4.64%
5	23,719	11.65%	32,446	19.12%	56,165	15.04%
6	128,189	62.95%	91,287	53.80%	219,476	58.79%
Total	203,631	100.00%	169,686	100.00%	373,317	100.00%

Figure 2.5.1.3

Proficiency Level: List 4–5 S602 Online



2.5.1.4 Grades 6–8

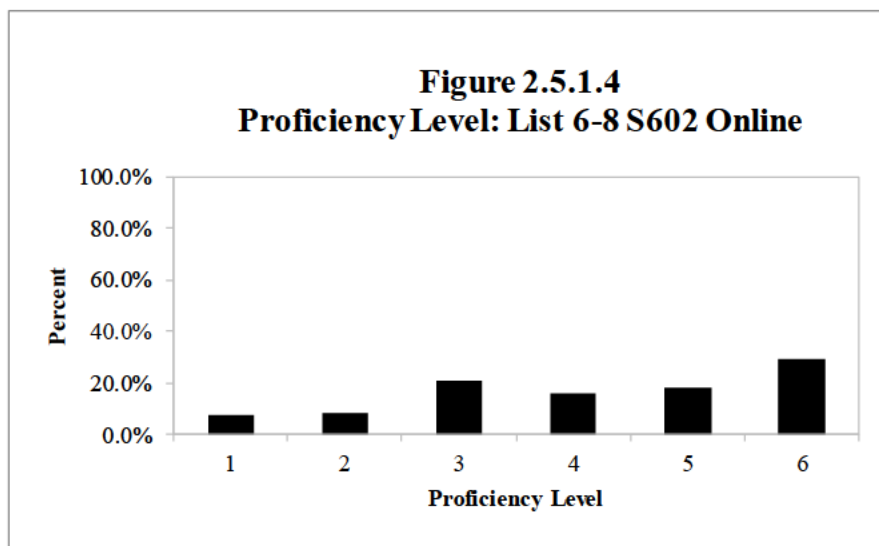
Table 2.5.1.4

Proficiency Level Distribution: List 6–8 S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	7,174	5.09%	10,716	7.26%	14,588	9.92%	32,478	7.46%
2	10,830	7.69%	12,503	8.47%	13,358	9.09%	36,691	8.42%
3	33,291	23.64%	29,929	20.26%	27,181	18.49%	90,401	20.76%
4	22,383	15.89%	24,142	16.35%	22,955	15.62%	69,480	15.95%
5	31,493	22.36%	29,420	19.92%	18,726	12.74%	79,639	18.29%
6	35,662	25.32%	40,983	27.75%	50,186	34.14%	126,831	29.12%
Total	140,833	100.00%	147,693	100.00%	146,994	100.00%	435,520	100.00%

Figure 2.5.1.4

Proficiency Level: List 6–8 S602 Online



2.5.1.5 Grade 9–12

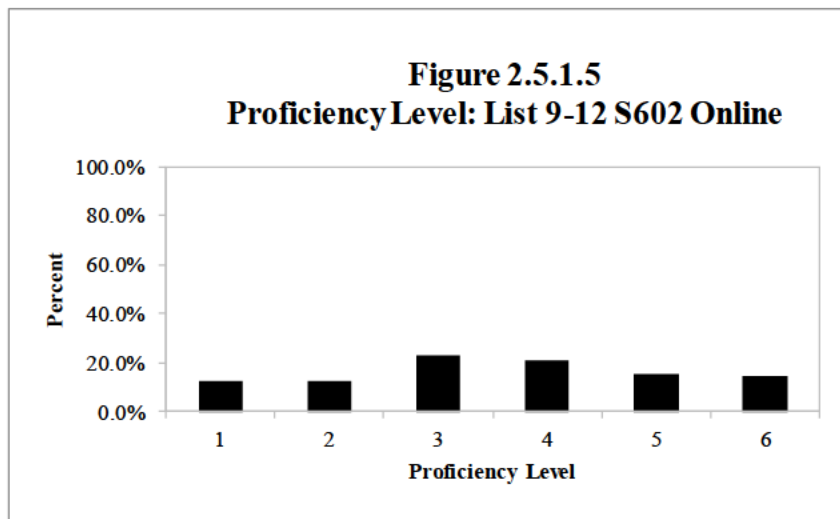
Table 2.5.1.5

Proficiency Level Distribution: List 9–12 S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	17,378	10.76%	17,061	12.23%	16,614	14.89%	12,430	15.37%	63,483	12.87%
2	20,474	12.68%	16,654	11.94%	14,275	12.79%	9,890	12.23%	61,293	12.42%
3	38,543	23.87%	33,987	24.37%	24,333	21.81%	18,539	22.93%	115,402	23.39%
4	35,122	21.76%	27,865	19.98%	22,238	19.93%	18,214	22.53%	103,439	20.97%
5	24,586	15.23%	22,774	16.33%	19,307	17.30%	10,659	13.18%	77,326	15.67%
6	25,334	15.69%	21,135	15.15%	14,808	13.27%	11,118	13.75%	72,395	14.67%
Total	161,437	100.00%	139,476	100.00%	111,575	100.00%	80,850	100.00%	493,338	100.00%

Figure 2.5.1.5

Proficiency Level: List 9–12 S602 Online



2.5.2 Reading

2.5.2.1 Grade 1

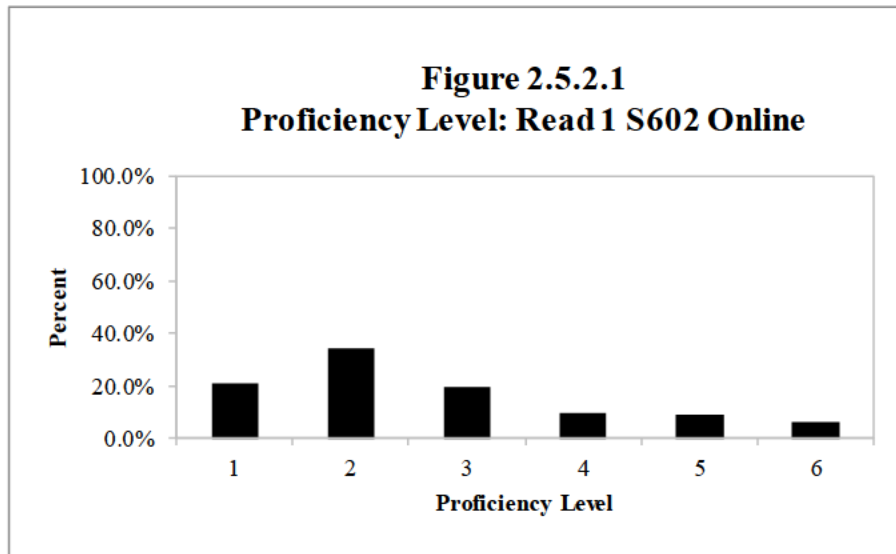
Table 2.5.2.1

Proficiency Level Distribution: Read 1 S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	46,553	20.87%	46,553	20.87%
2	76,336	34.22%	76,336	34.22%
3	42,961	19.26%	42,961	19.26%
4	21,927	9.83%	21,927	9.83%
5	20,813	9.33%	20,813	9.33%
6	14,511	6.50%	14,511	6.50%
Total	223,101	100.00%	223,101	100.00%

Figure 2.5.2.1

Proficiency Level: Read 1 S602 Online



2.5.2.2 Grades 2–3

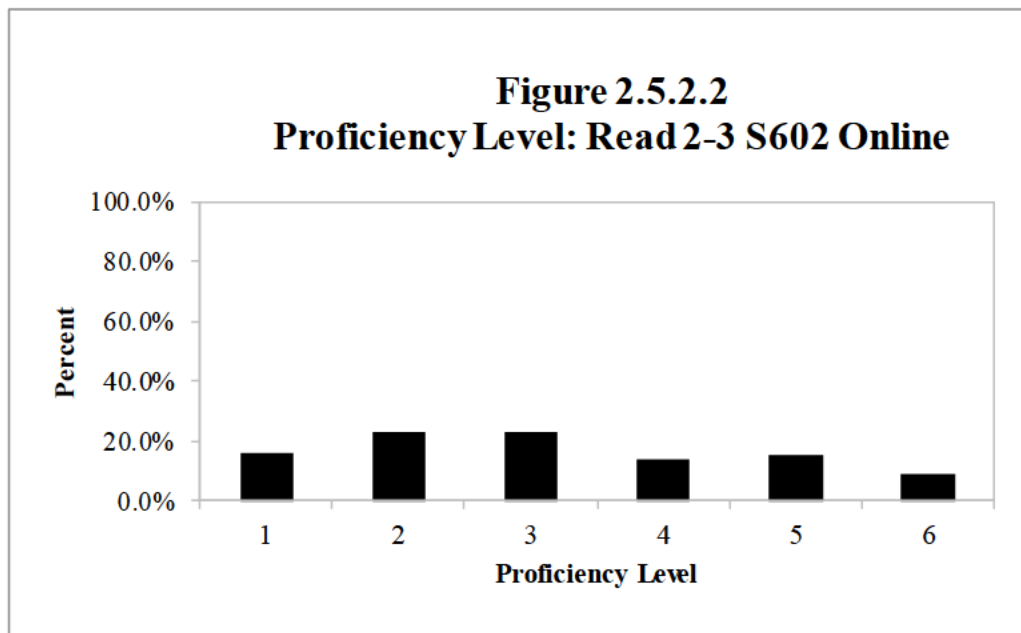
Table 2.5.2.2

Proficiency Level Distribution: Read 2–3 S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	28,092	12.17%	44,781	20.22%	72,873	16.12%
2	46,760	20.27%	57,020	25.75%	103,780	22.95%
3	64,938	28.14%	37,959	17.14%	102,897	22.76%
4	37,064	16.06%	26,812	12.11%	63,876	14.13%
5	38,089	16.51%	30,058	13.58%	68,147	15.07%
6	15,794	6.85%	24,789	11.20%	40,583	8.98%
Total	230,737	100.00%	221,419	100.00%	452,156	100.00%

Figure 2.5.2.2

Proficiency Level: Read 2–3 S602 Online



2.5.2.3 Grades 4–5

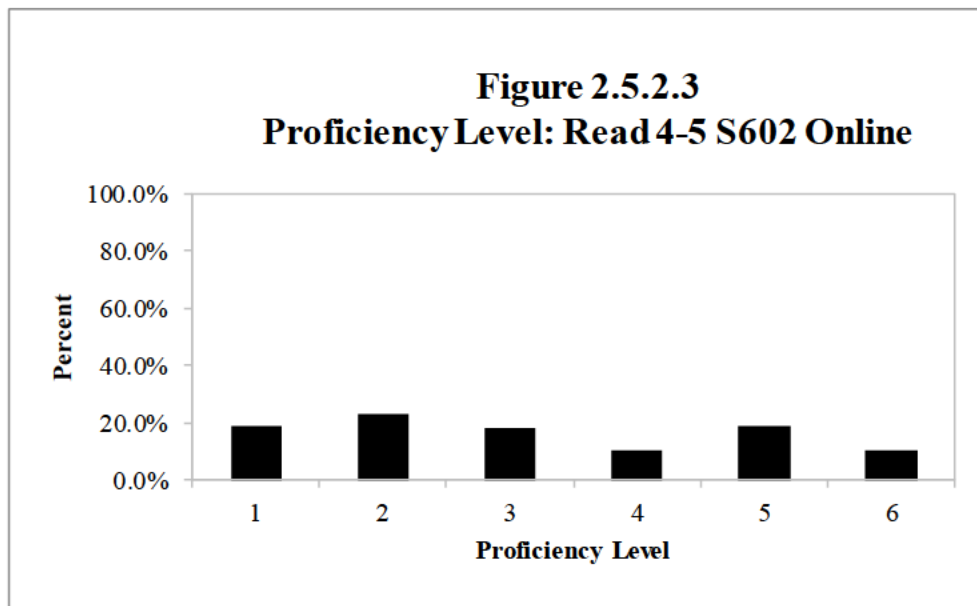
Table 2.5.2.3

Proficiency Level Distribution: Read 4–5 S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	33,593	16.45%	38,244	22.52%	71,837	19.20%
2	47,961	23.48%	37,755	22.23%	85,716	22.91%
3	36,062	17.65%	32,854	19.34%	68,916	18.42%
4	23,252	11.38%	15,807	9.31%	39,059	10.44%
5	40,792	19.97%	29,297	17.25%	70,089	18.73%
6	22,610	11.07%	15,894	9.36%	38,504	10.29%
Total	204,270	100.00%	169,851	100.00%	374,121	100.00%

Figure 2.5.2.3

Proficiency Level: Read 4–5 S602 Online



2.5.2.4 Grades 6–8

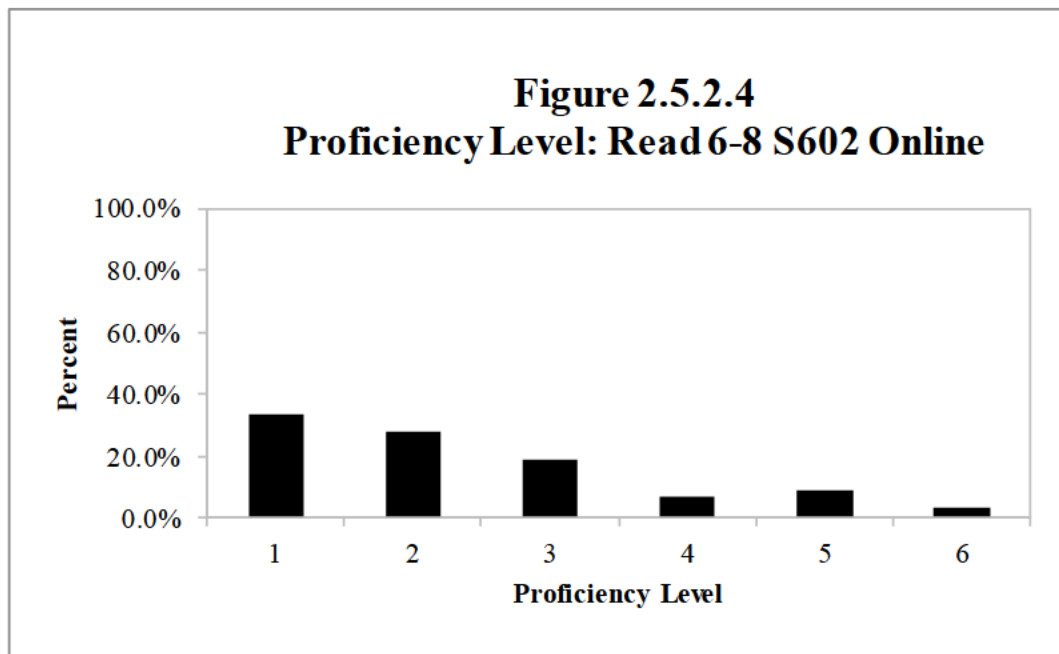
Table 2.5.2.4

Proficiency Level Distribution: Read 6–8 S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	48,763	33.52%	51,304	33.89%	51,108	34.16%	151,175	33.86%
2	46,159	31.73%	43,238	28.56%	35,683	23.85%	125,080	28.01%
3	26,476	18.20%	27,970	18.48%	30,501	20.39%	84,947	19.03%
4	8,248	5.67%	10,759	7.11%	10,602	7.09%	29,609	6.63%
5	11,936	8.20%	12,972	8.57%	14,682	9.81%	39,590	8.87%
6	3,911	2.69%	5,147	3.40%	7,026	4.70%	16,084	3.60%
Total	145,493	100.00%	151,390	100.00%	149,602	100.00%	446,485	100.00%

Figure 2.5.2.4

Proficiency Level: Read 6–8 S602 Online



2.5.2.5 Grades 9–12

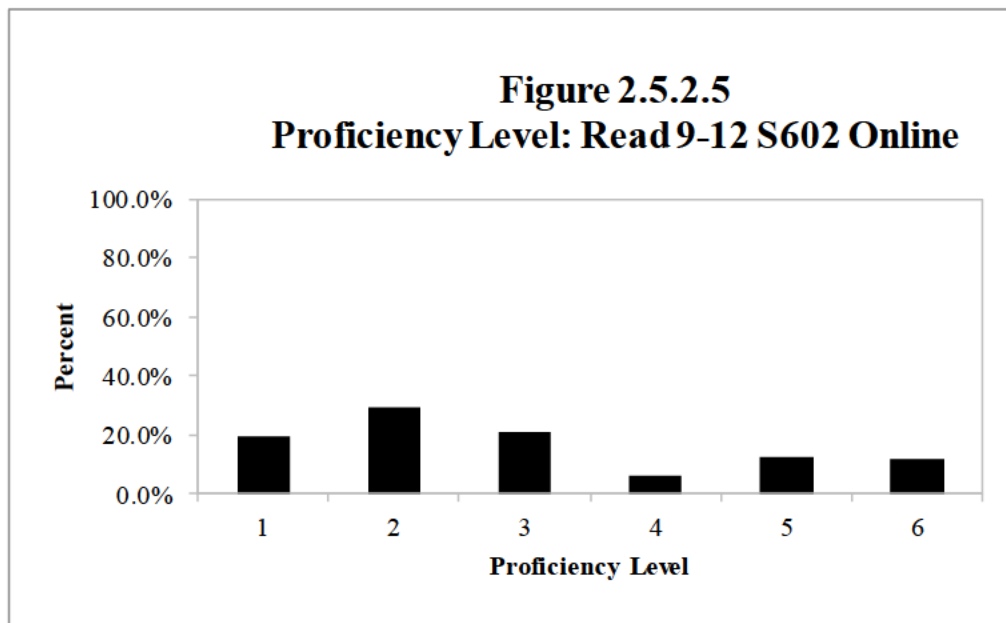
Table 2.5.2.5

Proficiency Level Distribution: Read 9–12 S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	29,821	18.54%	24,461	17.71%	23,722	21.50%	18,685	23.38%	96,689	19.76%
2	47,982	29.83%	41,647	30.16%	30,242	27.41%	22,976	28.75%	142,847	29.20%
3	35,512	22.07%	28,668	20.76%	21,356	19.36%	16,720	20.92%	102,256	20.90%
4	10,872	6.76%	9,019	6.53%	7,062	6.40%	3,338	4.18%	30,291	6.19%
5	19,791	12.30%	17,459	12.64%	13,736	12.45%	9,269	11.60%	60,255	12.32%
6	16,893	10.50%	16,845	12.20%	14,214	12.88%	8,935	11.18%	56,887	11.63%
Total	160,871	100.00%	138,099	100.00%	110,332	100.00%	79,923	100.00%	489,225	100.00%

Figure 2.5.2.5

Proficiency Level: Read 9–12 S602 Online



2.5.3 Writing

2.5.3.1 Grade 1

Table 2.5.3.1.1

Proficiency Level Distribution: Writ 1 A S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	103,439	49.35%	103,439	49.35%
2	75,148	35.85%	75,148	35.85%
3	30,952	14.77%	30,952	14.77%
4	54	0.03%	54	0.03%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	209,593	100.00%	209,593	100.00%

Figure 2.5.3.1.1

Proficiency Level: Writ 1 A S602 Online

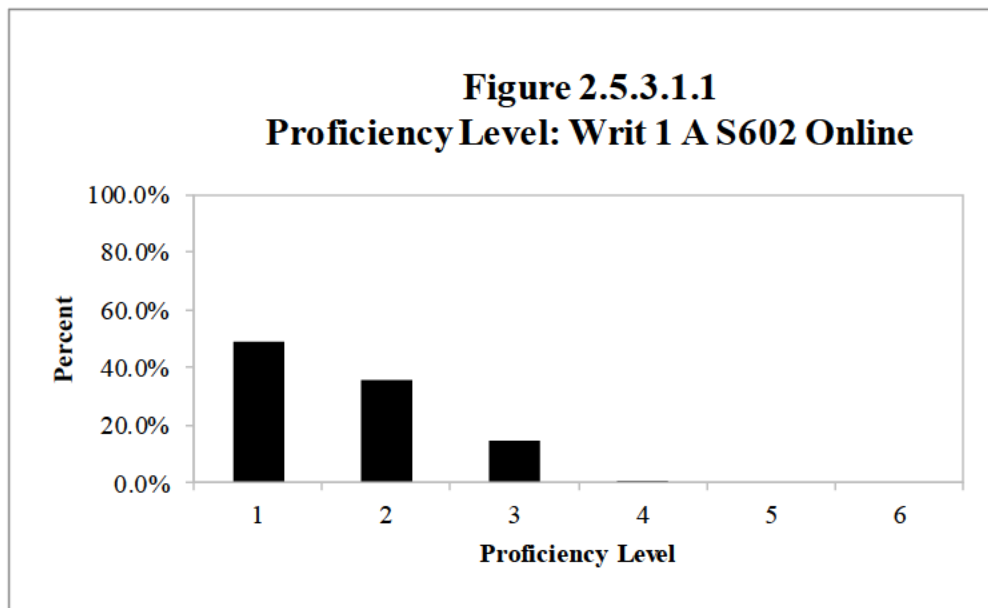


Table 2.5.3.1.2

Proficiency Level Distribution: Writ 1 B/C S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	687	2.63%	687	2.63%
2	5,794	22.16%	5,794	22.16%
3	17,305	66.20%	17,305	66.20%
4	2,257	8.63%	2,257	8.63%
5	77	0.29%	77	0.29%
6	21	0.08%	21	0.08%
Total	26,141	100.00%	26,141	100.00%

Figure 2.5.3.1.2

Proficiency Level: Writ 1 B/C S602 Online

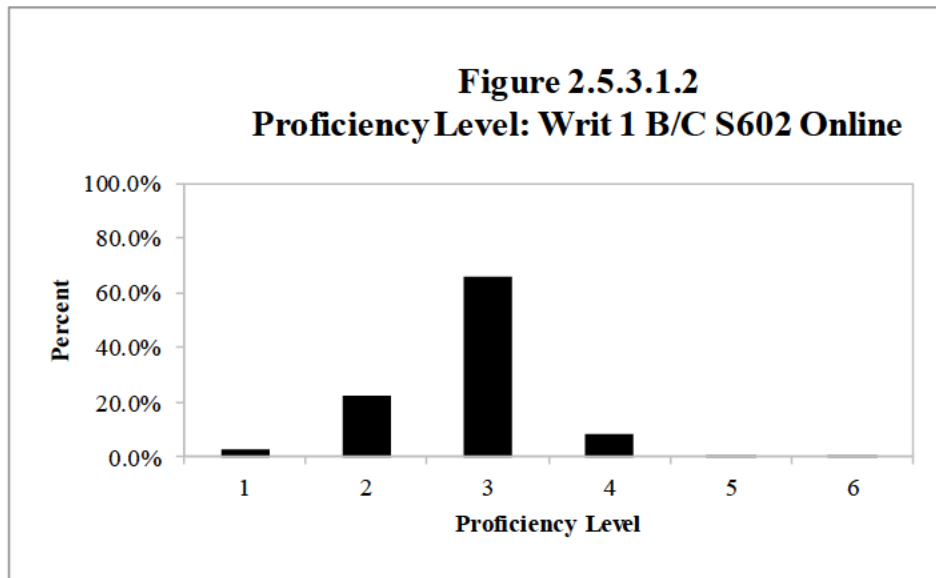


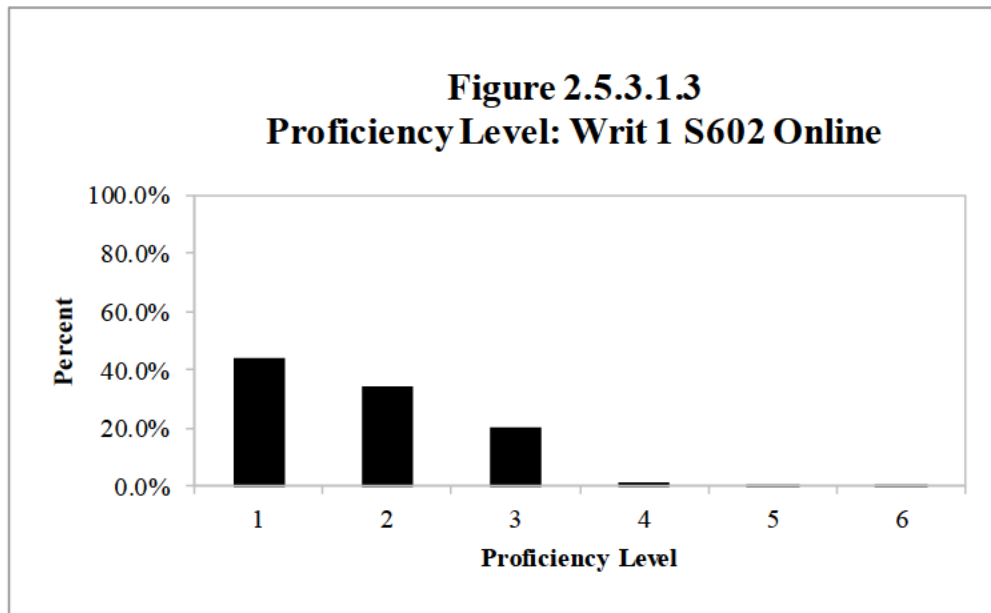
Table 2.5.3.1.3

Proficiency Level Distribution: Writ 1 S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	104,126	44.17%	104,126	44.17%
2	80,942	34.34%	80,942	34.34%
3	48,257	20.47%	48,257	20.47%
4	2,311	0.98%	2,311	0.98%
5	77	0.03%	77	0.03%
6	21	0.01%	21	0.01%
Total	235,734	100.00%	235,734	100.00%

Figure 2.5.3.1.3

Proficiency Level: Writ 1 S602 Online



2.5.3.2 Grades 2–3

Table 2.5.3.2.1

Proficiency Level Distribution: Writ 2–3 A S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	39,370	44.83%	32,512	44.34%	71,882	44.60%
2	35,842	40.81%	24,218	33.03%	60,060	37.27%
3	12,160	13.84%	16,361	22.31%	28,521	17.70%
4	458	0.52%	238	0.32%	696	0.43%
5	0	0.00%	1	0.00%	1	0.00%
6	0	0.00%	0	0.00%	0	0.00%
Total	87,830	100.00%	73,330	100.00%	161,160	100.00%

Figure 2.5.3.2.1

Proficiency Level: Writ 2–3 A S602 Online

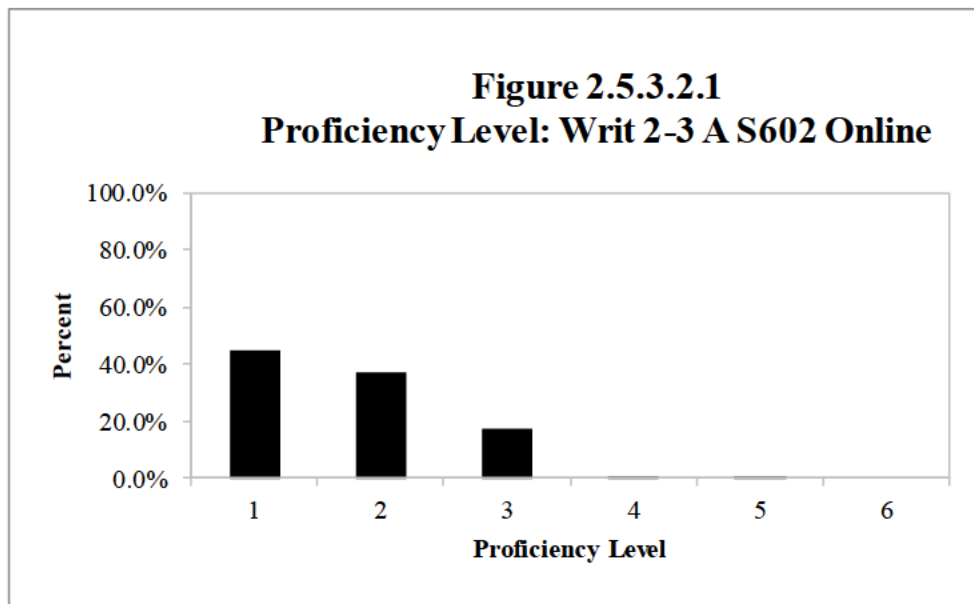


Table 2.5.3.2.2

Proficiency Level Distribution: Writ 2–3 B/C S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	10,441	6.55%	3,893	2.37%	14,334	4.43%
2	26,153	16.40%	9,746	5.93%	35,899	11.09%
3	111,244	69.78%	117,349	71.45%	228,593	70.63%
4	11,479	7.20%	32,431	19.75%	43,910	13.57%
5	102	0.06%	781	0.48%	883	0.27%
6	8	0.01%	35	0.02%	43	0.01%
Total	159,427	100.00%	164,235	100.00%	323,662	100.00%

Figure 2.5.3.2.2

Proficiency Level: Writ 2–3 B/C S602 Online

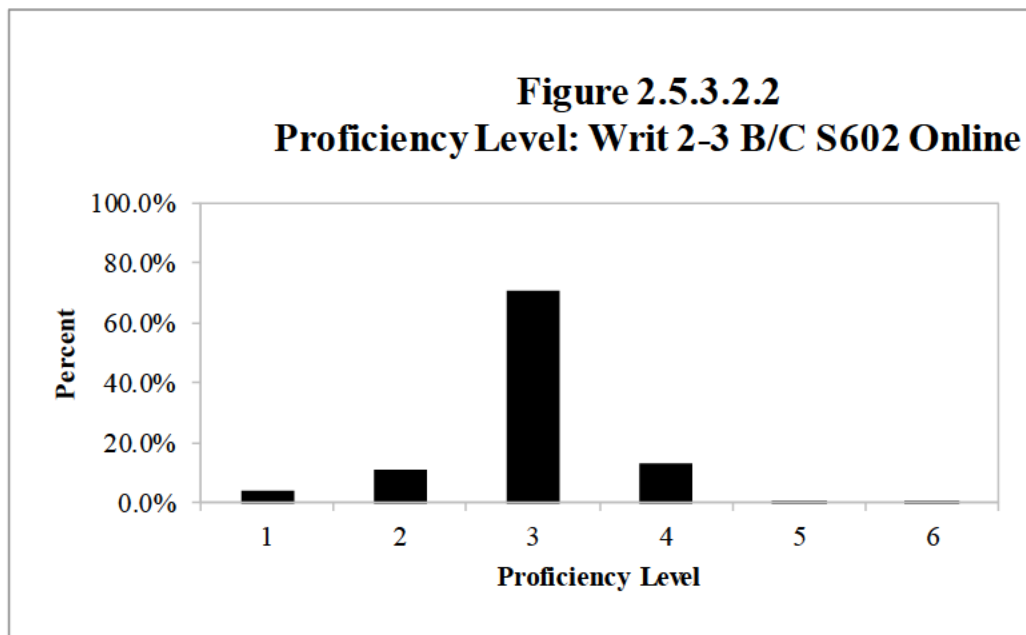


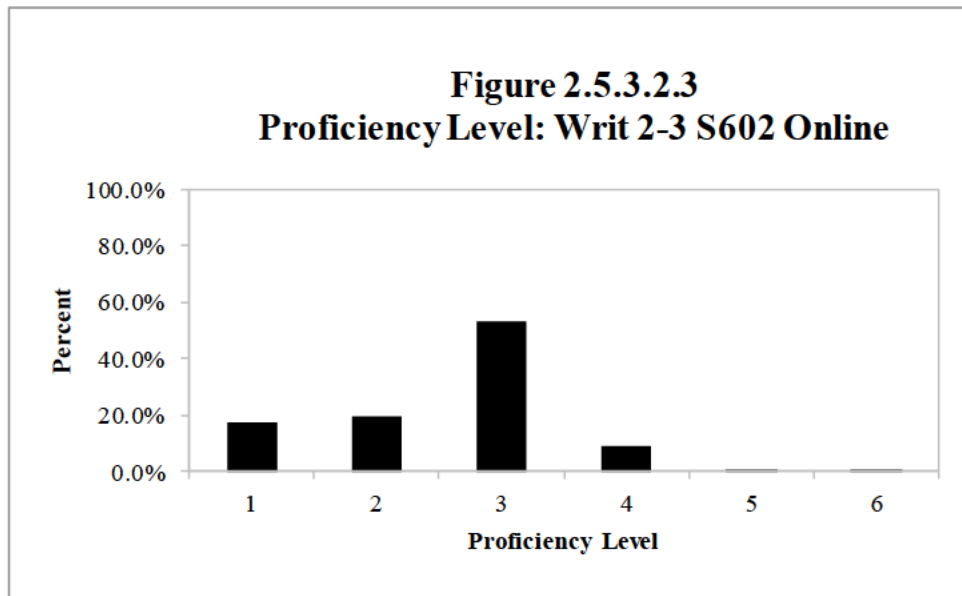
Table 2.5.3.2.3

Proficiency Level Distribution: Writ 2–3 S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	49,811	20.15%	36,405	15.32%	86,216	17.78%
2	61,995	25.07%	33,964	14.30%	95,959	19.79%
3	123,404	49.91%	133,710	56.28%	257,114	53.03%
4	11,937	4.83%	32,669	13.75%	44,606	9.20%
5	102	0.04%	782	0.33%	884	0.18%
6	8	0.00%	35	0.01%	43	0.01%
Total	247,257	100.00%	237,565	100.00%	484,822	100.00%

Figure 2.5.3.2.3

Proficiency Level: Writ 2–3 S602 Online



2.5.3.3 Grades 4–5

Table 2.5.3.3.1

Proficiency Level Distribution: Writ 4–5 A S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	24,962	44.54%	19,891	37.41%	44,853	41.07%
2	15,555	27.75%	13,776	25.91%	29,331	26.86%
3	15,130	27.00%	18,556	34.90%	33,686	30.84%
4	392	0.70%	947	1.78%	1,339	1.23%
5	5	0.01%	0	0.00%	5	0.00%
6	0	0.00%	0	0.00%	0	0.00%
Total	56,044	100.00%	53,170	100.00%	109,214	100.00%

Figure 2.5.3.3.1

Proficiency Level: Writ 4–5 A S602 Online

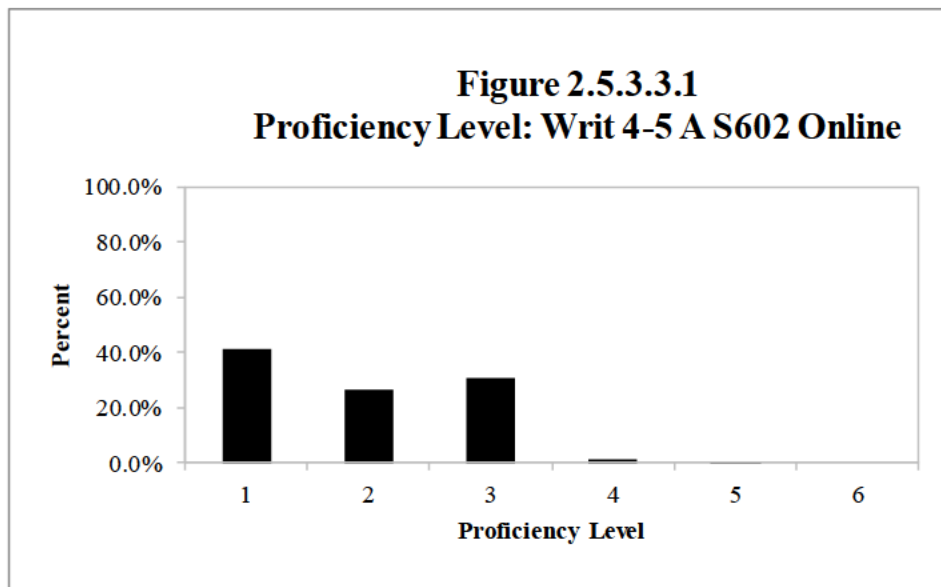


Table 2.5.3.3.2

Proficiency Level Distribution: Writ 4-5 B/C S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	2,422	1.59%	711	0.59%	3,133	1.15%
2	4,195	2.76%	3,147	2.61%	7,342	2.69%
3	91,440	60.16%	57,439	47.68%	148,879	54.64%
4	51,778	34.07%	55,824	46.34%	107,602	39.49%
5	1,691	1.11%	3,181	2.64%	4,872	1.79%
6	469	0.31%	176	0.15%	645	0.24%
Total	151,995	100.00%	120,478	100.00%	272,473	100.00%

Figure 2.5.3.3.2

Proficiency Level: Writ 4-5 B/C S602 Online

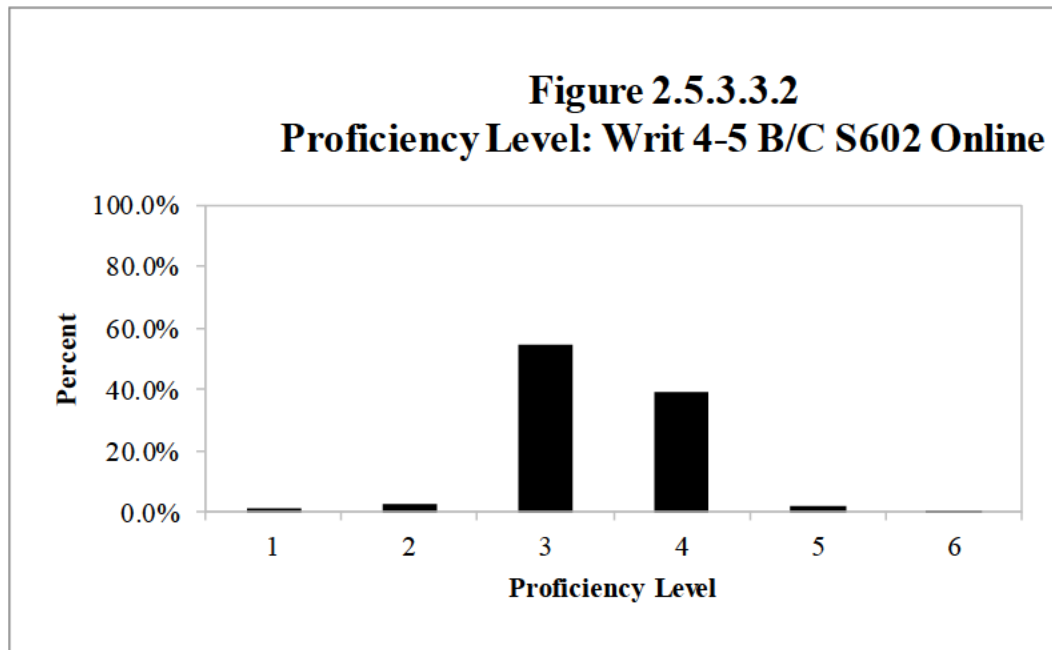


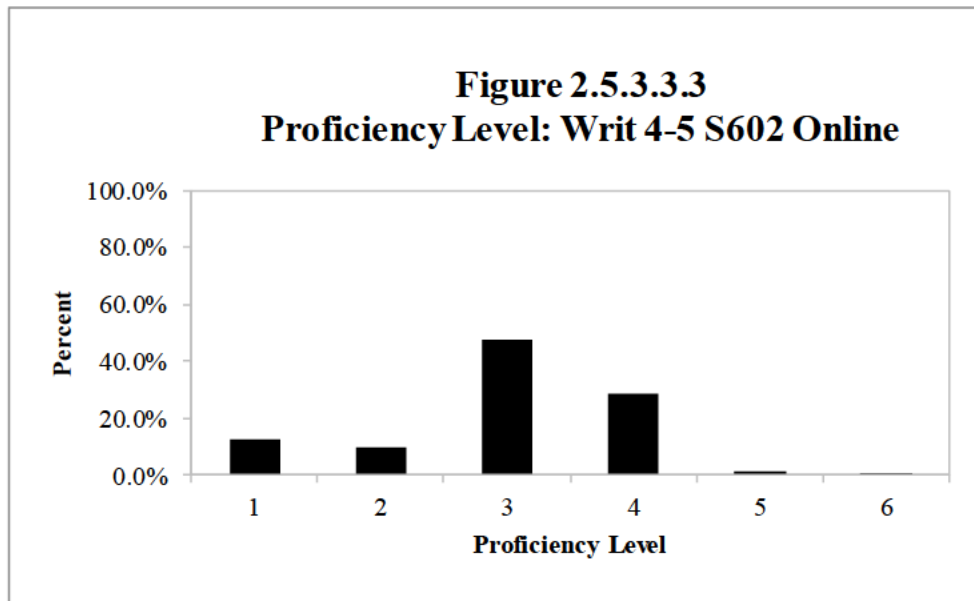
Table 2.5.3.3.3

Proficiency Level Distribution: Writ 4–5 S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	27,384	13.16%	20,602	11.86%	47,986	12.57%
2	19,750	9.49%	16,923	9.75%	36,673	9.61%
3	106,570	51.23%	75,995	43.76%	182,565	47.83%
4	52,170	25.08%	56,771	32.69%	108,941	28.54%
5	1,696	0.82%	3,181	1.83%	4,877	1.28%
6	469	0.23%	176	0.10%	645	0.17%
Total	208,039	100.00%	173,648	100.00%	381,687	100.00%

Figure 2.5.3.3.3

Proficiency Level: Writ 4–5 S602 Online



2.5.3.4 Grades 6–8

Table 2.5.3.4.1

Proficiency Level Distribution: Writ 6–8 A S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	25,917	42.45%	23,844	32.78%	28,900	39.48%	78,661	38.00%
2	16,522	27.06%	29,390	40.40%	19,851	27.12%	65,763	31.77%
3	18,175	29.77%	18,137	24.93%	23,690	32.37%	60,002	28.99%
4	443	0.73%	1,372	1.89%	750	1.02%	2,565	1.24%
5	0	0.00%	3	0.00%	3	0.00%	6	0.00%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	61,057	100.00%	72,746	100.00%	73,194	100.00%	206,997	100.00%

Figure 2.5.3.4.1

Proficiency Level: Writ 6–8 A S602 Online

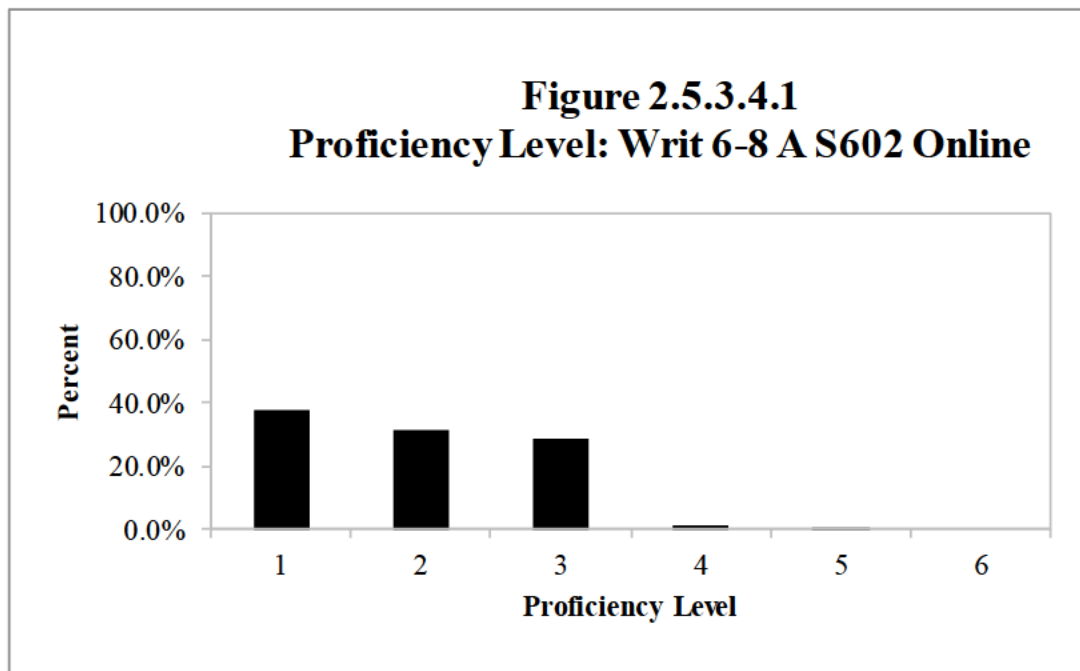


Table 2.5.3.4.2

Proficiency Level Distribution: Writ 6–8 B/C S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	1,298	1.41%	615	0.71%	639	0.76%	2,552	0.97%
2	10,473	11.40%	9,658	11.13%	4,908	5.88%	25,039	9.55%
3	67,125	73.09%	52,059	60.01%	61,291	73.37%	180,475	68.85%
4	12,836	13.98%	24,331	28.05%	16,445	19.69%	53,612	20.45%
5	112	0.12%	93	0.11%	247	0.30%	452	0.17%
6	0	0.00%	0	0.00%	6	0.01%	6	0.00%
Total	91,844	100.00%	86,756	100.00%	83,536	100.00%	262,136	100.00%

Figure 2.5.3.4.2

Proficiency Level: Writ 6–8 B/C S602 Online

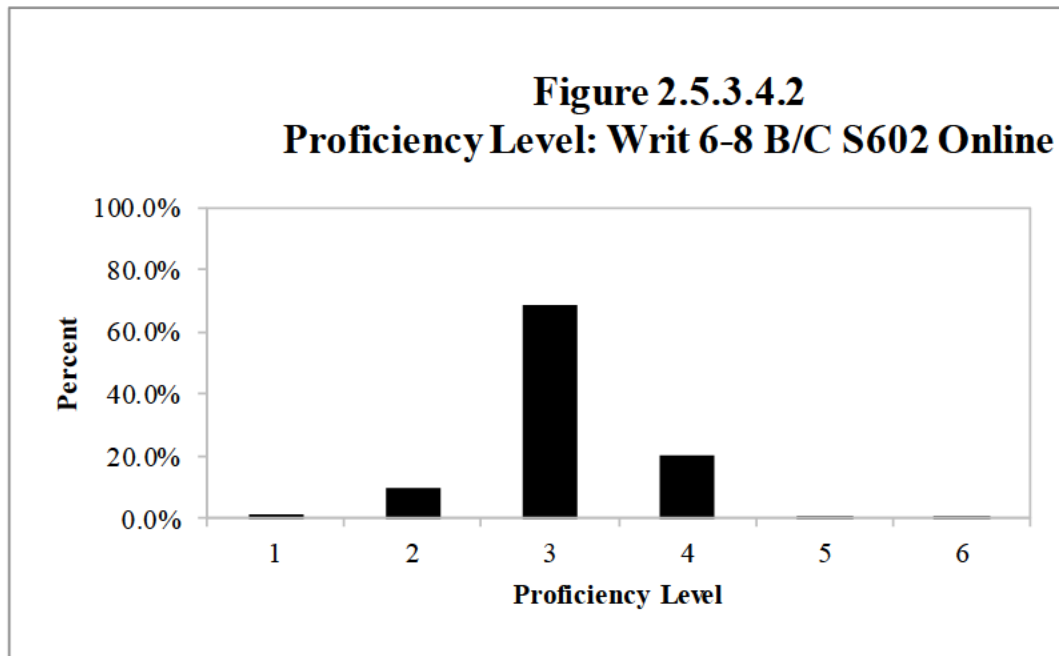


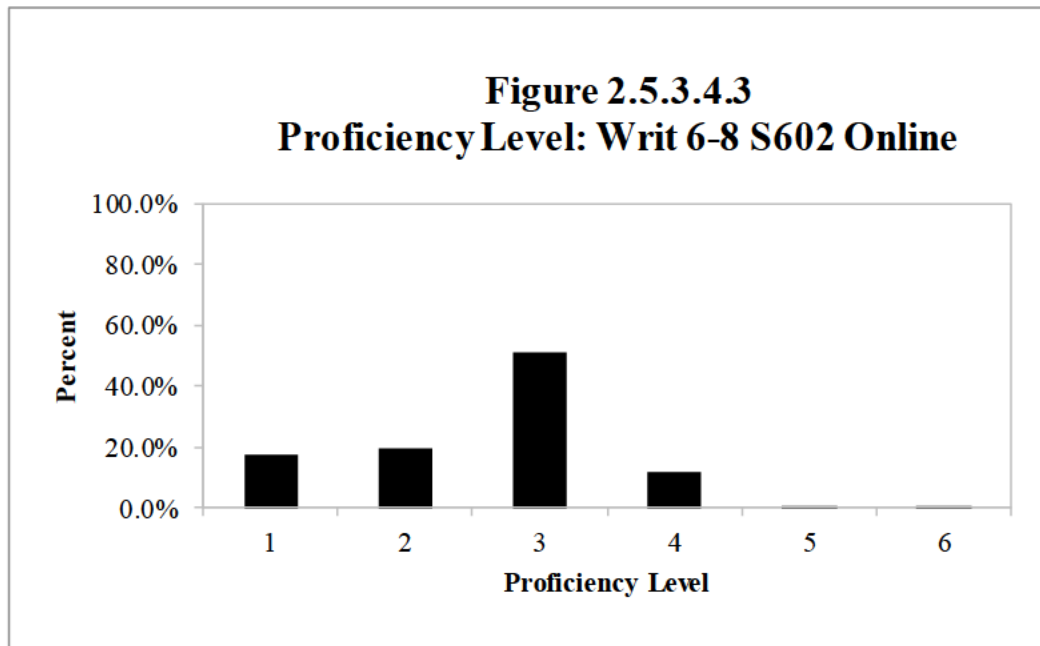
Table 2.5.3.4.3

Proficiency Level Distribution: Writ 6–8 S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	27,215	17.80%	24,459	15.33%	29,539	18.85%	81,213	17.31%
2	26,995	17.66%	39,048	24.48%	24,759	15.80%	90,802	19.36%
3	85,300	55.79%	70,196	44.01%	84,981	54.22%	240,477	51.26%
4	13,279	8.68%	25,703	16.11%	17,195	10.97%	56,177	11.97%
5	112	0.07%	96	0.06%	250	0.16%	458	0.10%
6	0	0.00%	0	0.00%	6	0.00%	6	0.00%
Total	152,901	100.00%	159,502	100.00%	156,730	100.00%	469,133	100.00%

Figure 2.5.3.4.3

Proficiency Level: Writ 6–8 S602 Online



2.5.3.5 Grades 9–12

Table 2.5.3.5.1

Proficiency Level Distribution: Writ 9–12 A S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	21,242	30.32%	14,687	27.89%	16,391	39.06%	13,256	48.51%	65,576	34.15%
2	23,617	33.72%	16,537	31.40%	11,647	27.75%	4,228	15.47%	56,029	29.18%
3	19,855	28.34%	19,518	37.06%	12,002	28.60%	8,477	31.02%	59,852	31.17%
4	5,295	7.56%	1,868	3.55%	1,920	4.57%	1,355	4.96%	10,438	5.44%
5	38	0.05%	52	0.10%	9	0.02%	9	0.03%	108	0.06%
6	1	0.00%	0	0.00%	0	0.00%	0	0.00%	1	0.00%
Total	70,048	100.00%	52,662	100.00%	41,969	100.00%	27,325	100.00%	192,004	100.00%

Figure 2.5.3.5.1

Proficiency Level: Writ 9–12 A S602 Online

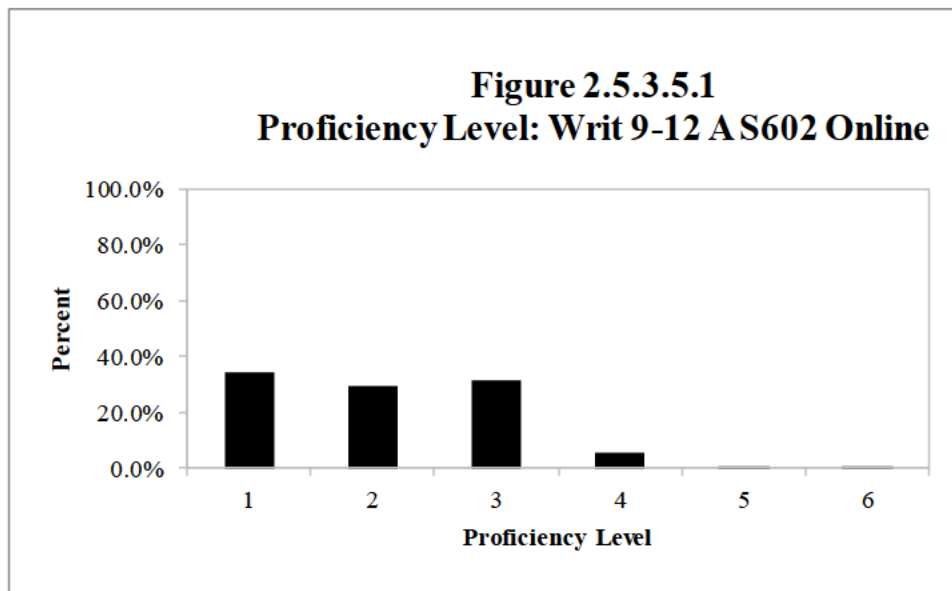


Table 2.5.3.5.2

Proficiency Level Distribution: Writ 9–12 B/C S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	2,797	2.84%	3,383	3.68%	4,386	5.98%	6,811	12.14%	17,377	5.43%
2	12,434	12.61%	9,037	9.83%	12,521	17.07%	6,996	12.47%	40,988	12.81%
3	53,953	54.73%	64,726	70.44%	42,769	58.31%	32,523	57.95%	193,971	60.63%
4	28,906	29.32%	14,184	15.44%	13,533	18.45%	9,710	17.30%	66,333	20.73%
5	472	0.48%	557	0.61%	142	0.19%	83	0.15%	1,254	0.39%
6	12	0.01%	1	0.00%	0	0.00%	0	0.00%	13	0.00%
Total	98,574	100.00%	91,888	100.00%	73,351	100.00%	56,123	100.00%	319,936	100.00%

Figure 2.5.3.5.2

Proficiency Level: Writ 9–12 B/C S602 Online

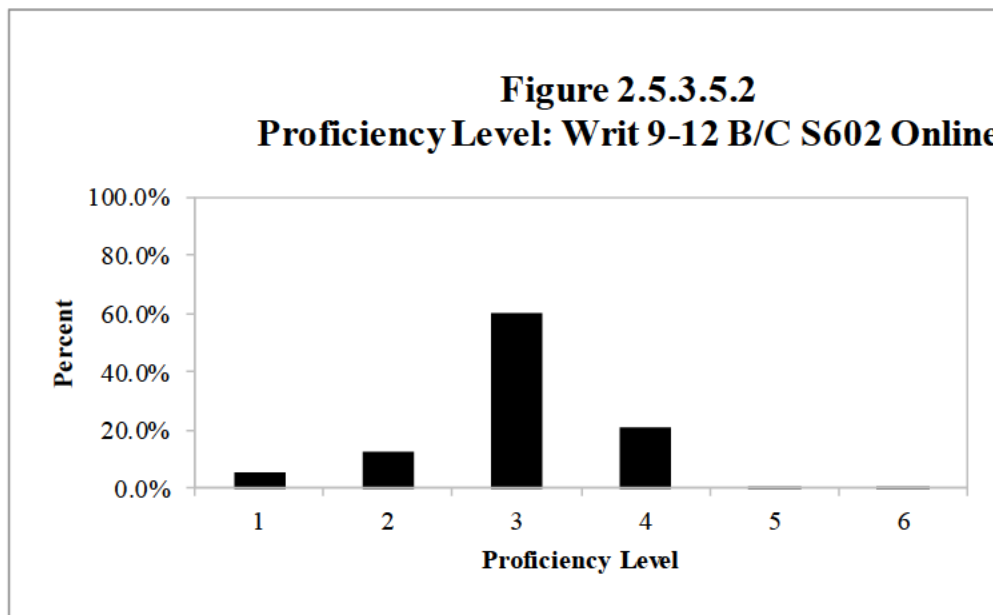


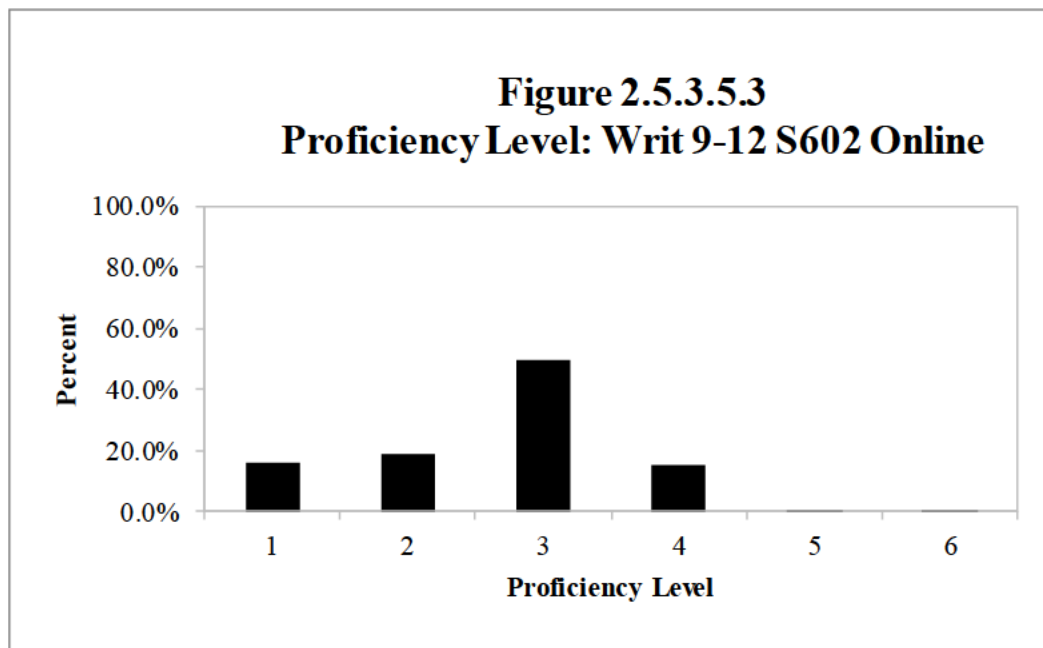
Table 2.5.3.5.3

Proficiency Level Distribution: Writ 9–12 S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	24,039	14.26%	18,070	12.50%	20,777	18.02%	20,067	24.05%	82,953	16.20%
2	36,051	21.38%	25,574	17.69%	24,168	20.96%	11,224	13.45%	97,017	18.95%
3	73,808	43.77%	84,244	58.28%	54,771	47.49%	41,000	49.13%	253,823	49.58%
4	34,201	20.28%	16,052	11.10%	15,453	13.40%	11,065	13.26%	76,771	15.00%
5	510	0.30%	609	0.42%	151	0.13%	92	0.11%	1,362	0.27%
6	13	0.01%	1	0.00%	0	0.00%	0	0.00%	14	0.00%
Total	168,622	100.00%	144,550	100.00%	115,320	100.00%	83,448	100.00%	511,940	100.00%

Figure 2.5.3.5.3

Proficiency Level: Writ 9–12 S602 Online



2.5.4 Speaking

2.5.4.1 Grade 1

Table 2.5.4.1.1

Proficiency Level Distribution: Spek 1 Pre-A S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	14,805	100.00%	14,805	100.00%
Total	14,805	100.00%	14,805	100.00%

Figure 2.5.4.1.1

Proficiency Level: Spek 1 Pre-A S602 Online

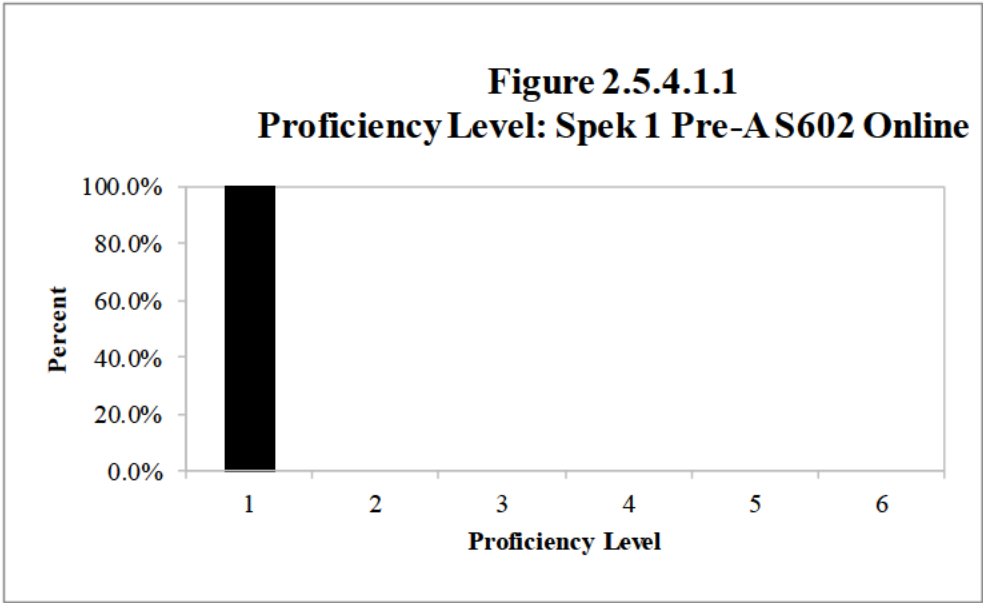


Table 2.5.4.1.2

Proficiency Level Distribution: Spek 1 A S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	44,493	40.97%	44,493	40.97%
2	35,656	32.84%	35,656	32.84%
3	25,786	23.75%	25,786	23.75%
4	2,524	2.32%	2,524	2.32%
5	132	0.12%	132	0.12%
6	0	0.00%	0	0.00%
Total	108,591	100.00%	108,591	100.00%

Figure 2.5.4.1.2

Proficiency Level: Spek 1 A S602 Online

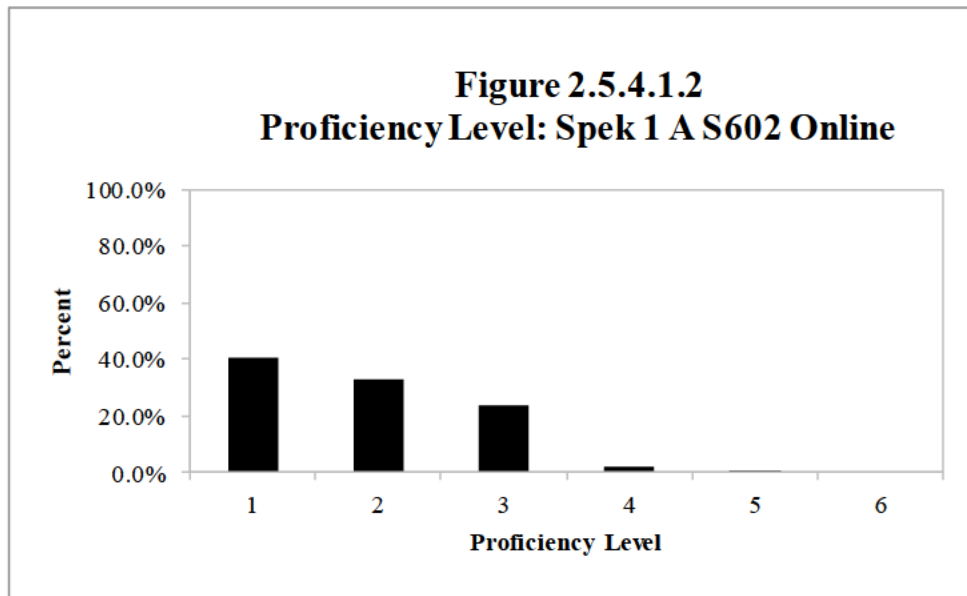


Table 2.5.4.1.3

Proficiency Level Distribution: Spek 1 B/C S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	6,389	7.02%	6,389	7.02%
2	29,776	32.72%	29,776	32.72%
3	42,277	46.45%	42,277	46.45%
4	11,324	12.44%	11,324	12.44%
5	1,132	1.24%	1,132	1.24%
6	111	0.12%	111	0.12%
Total	91,009	100.00%	91,009	100.00%

Figure 2.5.4.1.3

Proficiency Level: Spek 1 B/C S602 Online

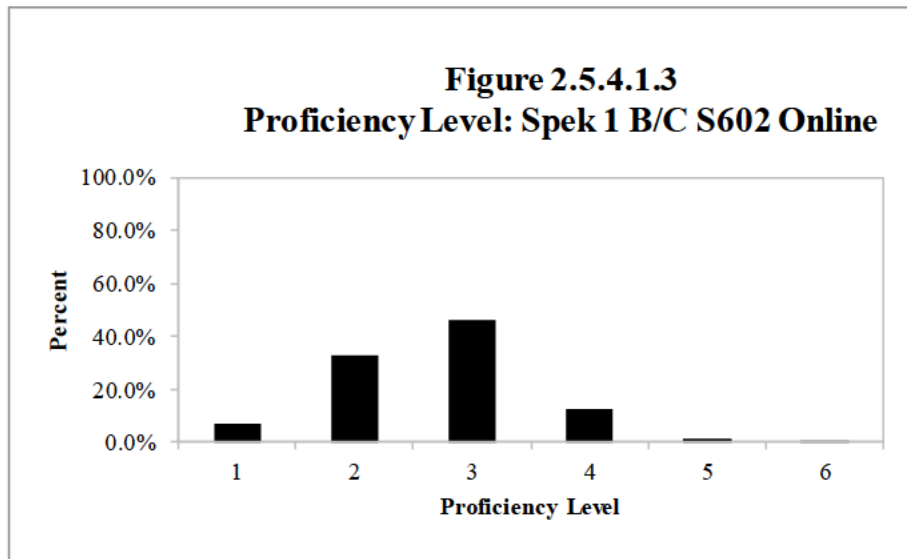


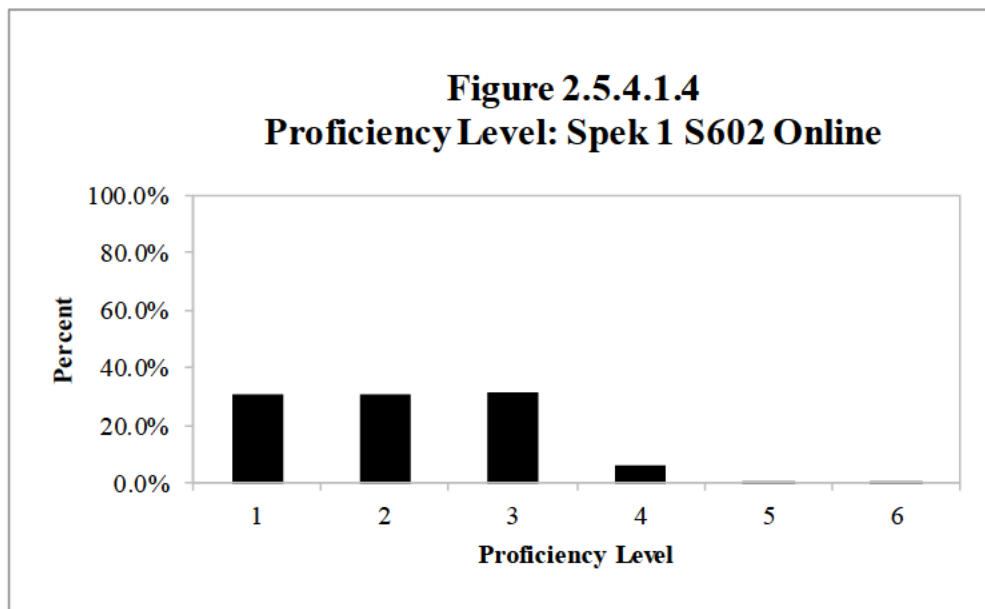
Table 2.5.4.1.4

Proficiency Level Distribution: Spek 1 S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	65,687	30.64%	65,687	30.64%
2	65,432	30.52%	65,432	30.52%
3	68,063	31.75%	68,063	31.75%
4	13,848	6.46%	13,848	6.46%
5	1,264	0.59%	1,264	0.59%
6	111	0.05%	111	0.05%
Total	214,405	100.00%	214,405	100.00%

Figure 2.5.4.1.4

Proficiency Level: Spek 1 S602 Online



2.5.4.2 Grades 2–3

Table 2.5.4.2.1

Proficiency Level Distribution: Spek 2–3 Pre-A S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	9,393	100.00%	20,574	100.00%	29,967	100.00%
Total	9,393	100.00%	20,574	100.00%	29,967	100.00%

Figure 2.5.4.2.1

Proficiency Level: Spek 2–3 Pre-A S602 Online

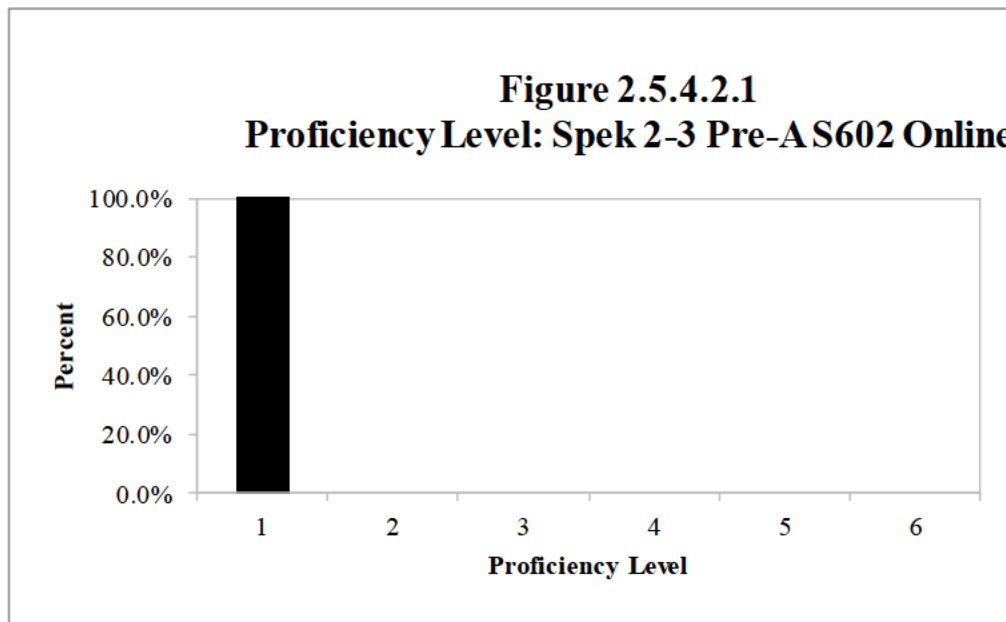


Table 2.5.4.2.2

Proficiency Level Distribution: Spek 2-3 A S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	34,847	46.67%	19,612	28.98%	54,459	38.26%
2	22,998	30.80%	21,674	32.03%	44,672	31.39%
3	13,681	18.32%	24,038	35.53%	37,719	26.50%
4	3,109	4.16%	2,240	3.31%	5,349	3.76%
5	24	0.03%	101	0.15%	125	0.09%
6	0	0.00%	0	0.00%	0	0.00%
Total	74,659	100.00%	67,665	100.00%	142,324	100.00%

Figure 2.5.4.2.2

Proficiency Level: Spek 2-3 A S602 Online

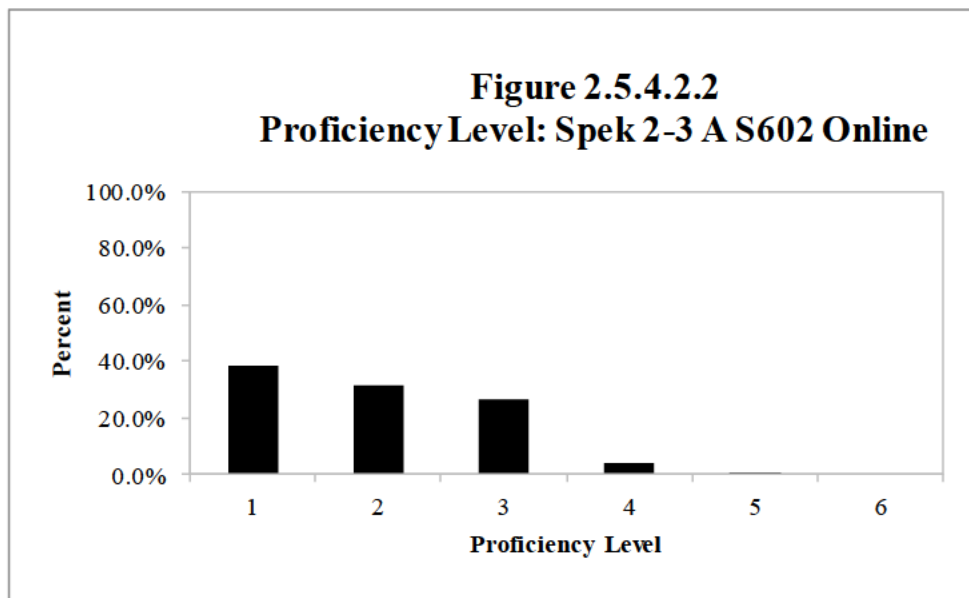


Table 2.5.4.2.3

Proficiency Level Distribution: Spek 2–3 B/C S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	11,690	8.26%	6,143	4.66%	17,833	6.53%
2	49,669	35.09%	25,257	19.17%	74,926	27.42%
3	55,635	39.30%	71,198	54.04%	126,833	46.41%
4	21,910	15.48%	26,112	19.82%	48,022	17.57%
5	2,502	1.77%	2,642	2.01%	5,144	1.88%
6	149	0.11%	393	0.30%	542	0.20%
Total	141,555	100.00%	131,745	100.00%	273,300	100.00%

Figure 2.5.4.2.3

Proficiency Level: Spek 2–3 B/C S602 Online

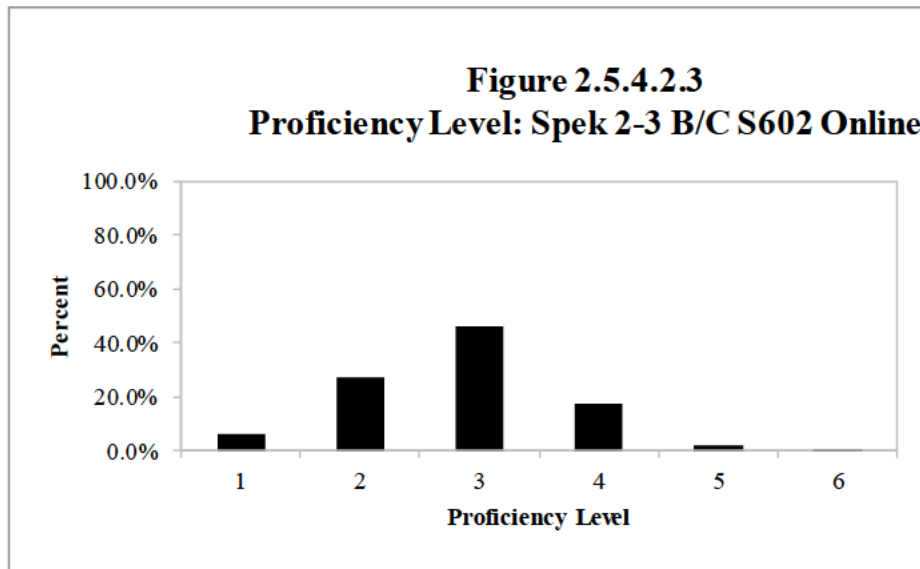


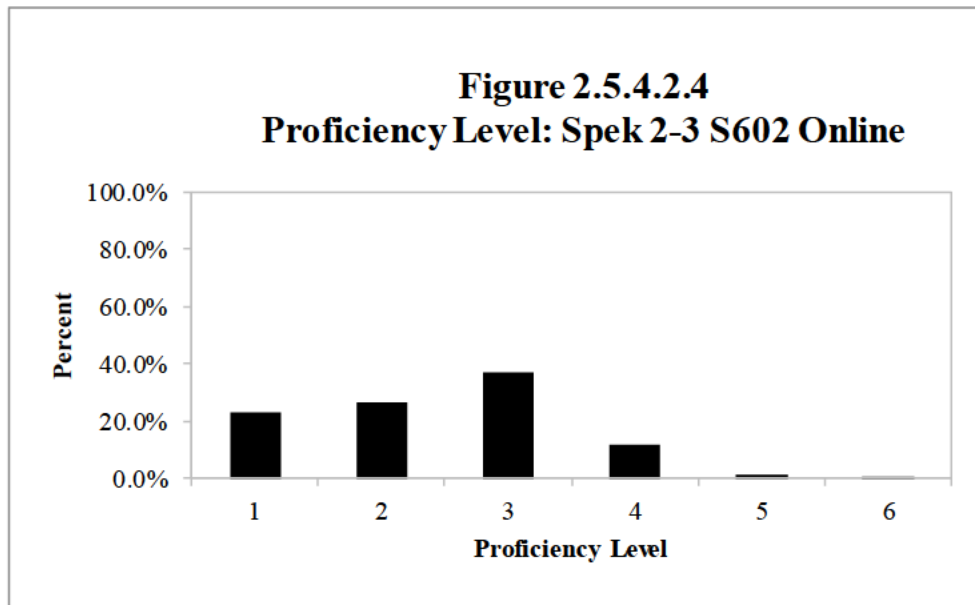
Table 2.5.4.2.4

Proficiency Level Distribution: Spek 2–3 S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	55,930	24.79%	46,329	21.06%	102,259	22.95%
2	72,667	32.21%	46,931	21.33%	119,598	26.84%
3	69,316	30.72%	95,236	43.29%	164,552	36.93%
4	25,019	11.09%	28,352	12.89%	53,371	11.98%
5	2,526	1.12%	2,743	1.25%	5,269	1.18%
6	149	0.07%	393	0.18%	542	0.12%
Total	225,607	100.00%	219,984	100.00%	445,591	100.00%

Figure 2.5.4.2.4

Proficiency Level: Spek 2–3 S602 Online



2.5.4.3 Grades 4–5

Table 2.5.4.3.1

Proficiency Level Distribution: Spek 4–5 Pre-A S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	4,046	100.00%	8,610	100.00%	12,656	100.00%
Total	4,046	100.00%	8,610	100.00%	12,656	100.00%

Figure 2.5.4.3.1

Proficiency Level: Spek 4–5 Pre-A S602 Online

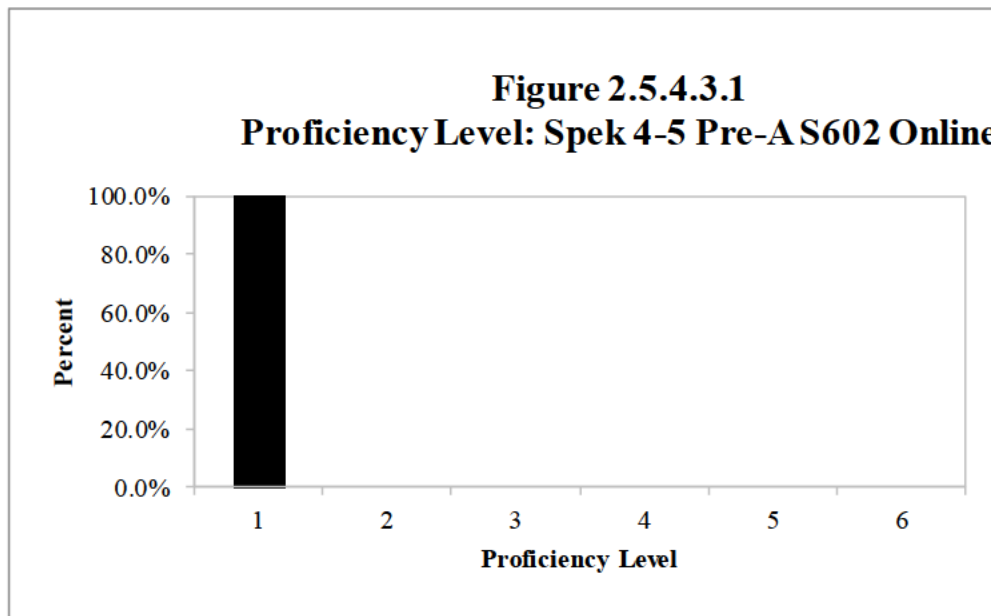


Table 2.5.4.3.2

Proficiency Level Distribution: Spek 4-5 A S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	22,666	52.07%	20,918	63.83%	43,584	57.12%
2	11,177	25.68%	7,711	23.53%	18,888	24.75%
3	7,366	16.92%	3,169	9.67%	10,535	13.81%
4	2,233	5.13%	825	2.52%	3,058	4.01%
5	88	0.20%	149	0.45%	237	0.31%
6	0	0.00%	0	0.00%	0	0.00%
Total	43,530	100.00%	32,772	100.00%	76,302	100.00%

Figure 2.5.4.3.2

Proficiency Level: Spek 4-5 A S602 Online

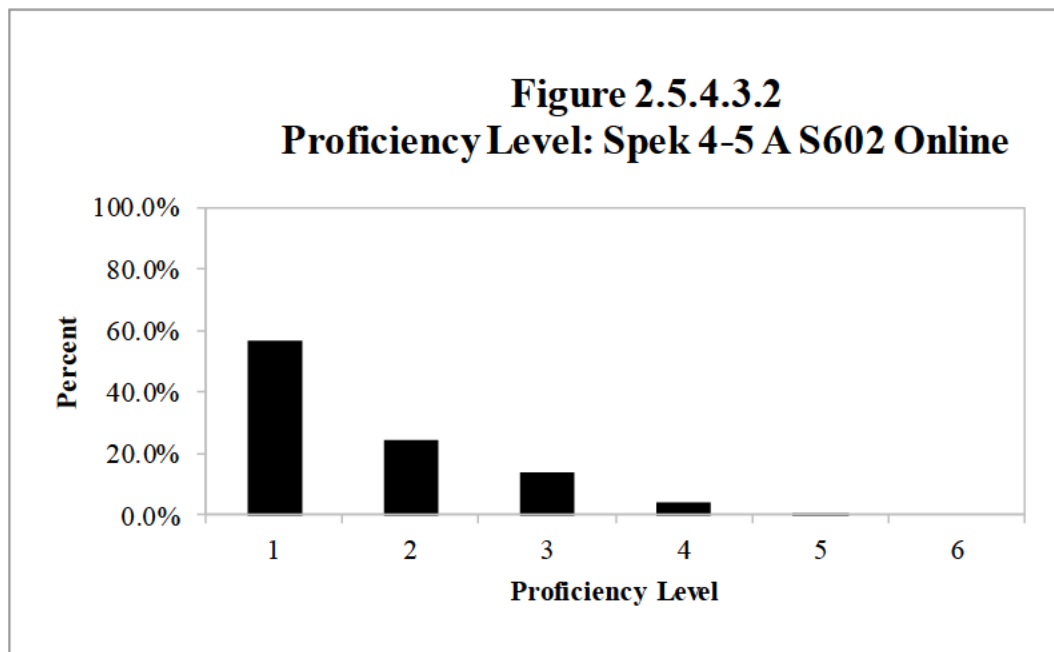


Table 2.5.4.3.3

Proficiency Level Distribution: Spek 4-5 B/C S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	4,387	2.77%	6,197	4.76%	10,584	3.67%
2	37,109	23.40%	26,443	20.31%	63,552	22.01%
3	58,646	36.98%	48,248	37.06%	106,894	37.02%
4	49,167	31.01%	41,145	31.60%	90,312	31.27%
5	8,094	5.10%	7,116	5.47%	15,210	5.27%
6	1,166	0.74%	1,053	0.81%	2,219	0.77%
Total	158,569	100.00%	130,202	100.00%	288,771	100.00%

Figure 2.5.4.3.3

Proficiency Level: Spek 4-5 B/C S602 Online

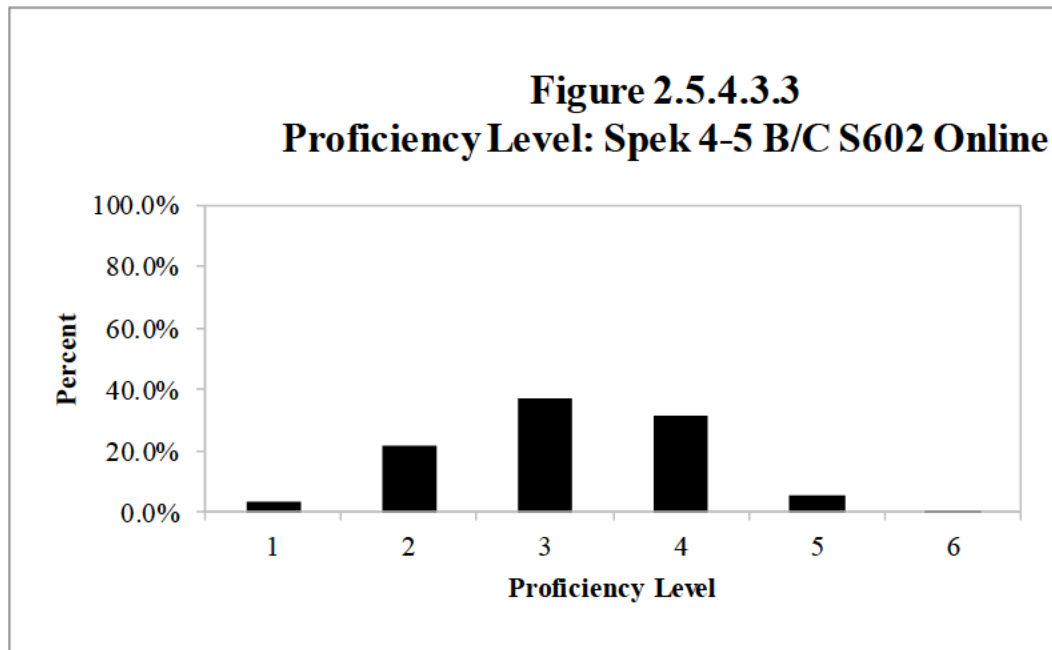


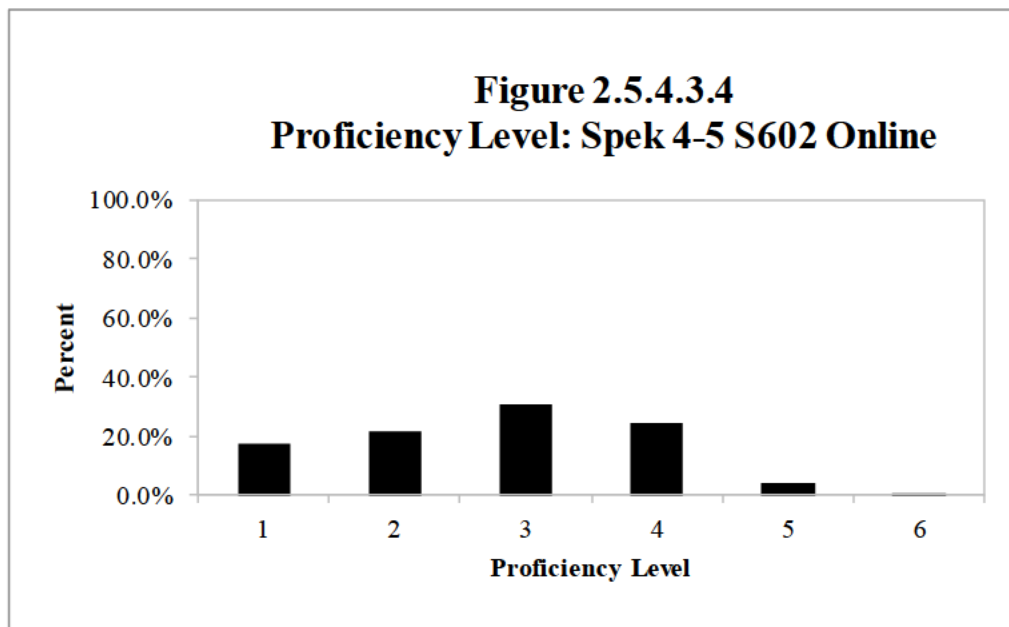
Table 2.5.4.3.4

Proficiency Level Distribution: Spek 4-5 S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	31,099	15.09%	35,725	20.82%	66,824	17.69%
2	48,286	23.42%	34,154	19.91%	82,440	21.83%
3	66,012	32.02%	51,417	29.97%	117,429	31.09%
4	51,400	24.93%	41,970	24.46%	93,370	24.72%
5	8,182	3.97%	7,265	4.23%	15,447	4.09%
6	1,166	0.57%	1,053	0.61%	2,219	0.59%
Total	206,145	100.00%	171,584	100.00%	377,729	100.00%

Figure 2.5.4.3.4

Proficiency Level: Spek 4-5 S602 Online



2.5.4.4 Grades 6–8

Table 2.5.4.4.1

Proficiency Level Distribution: Spek 6–8 Pre-A S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	6,152	100.00%	10,204	100.00%	14,536	100.00%	30,892	100.00%
Total	6,152	100.00%	10,204	100.00%	14,536	100.00%	30,892	100.00%

Figure 2.5.4.4.1

Proficiency Level: Spek 6–8 Pre-A S602 Online

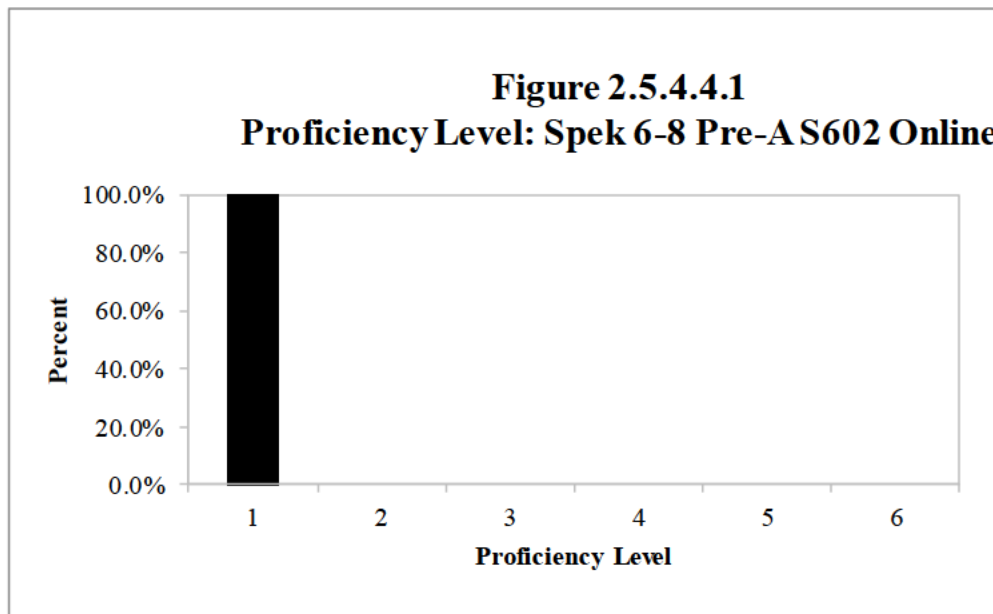


Table 2.5.4.4.2

Proficiency Level Distribution: Spek 6–8 A S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	20,814	59.32%	19,286	75.06%	25,023	53.19%	65,123	60.40%
2	9,514	27.11%	3,275	12.75%	8,751	18.60%	21,540	19.98%
3	3,998	11.39%	2,662	10.36%	12,164	25.86%	18,824	17.46%
4	756	2.15%	467	1.82%	1,104	2.35%	2,327	2.16%
5	7	0.02%	4	0.02%	0	0.00%	11	0.01%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	35,089	100.00%	25,694	100.00%	47,042	100.00%	107,825	100.00%

Figure 2.5.4.4.2

Proficiency Level: Spek 6–8 A S602 Online

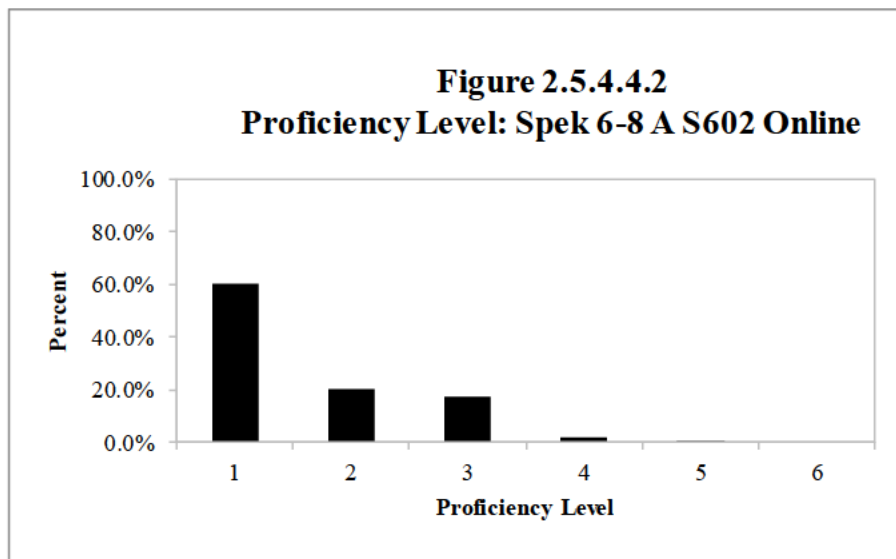


Table 2.5.4.4.3

Proficiency Level Distribution: Spek 6–8 B/C S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	13,530	12.94%	20,046	17.37%	8,852	9.90%	42,428	13.71%
2	20,478	19.59%	25,072	21.72%	14,460	16.17%	60,010	19.40%
3	52,025	49.77%	44,407	38.47%	44,884	50.20%	141,316	45.68%
4	17,924	17.15%	24,699	21.40%	20,523	22.95%	63,146	20.41%
5	534	0.51%	1,107	0.96%	505	0.56%	2,146	0.69%
6	35	0.03%	93	0.08%	186	0.21%	314	0.10%
Total	104,526	100.00%	115,424	100.00%	89,410	100.00%	309,360	100.00%

Figure 2.5.4.4.3

Proficiency Level: Spek 6–8 B/C S602 Online

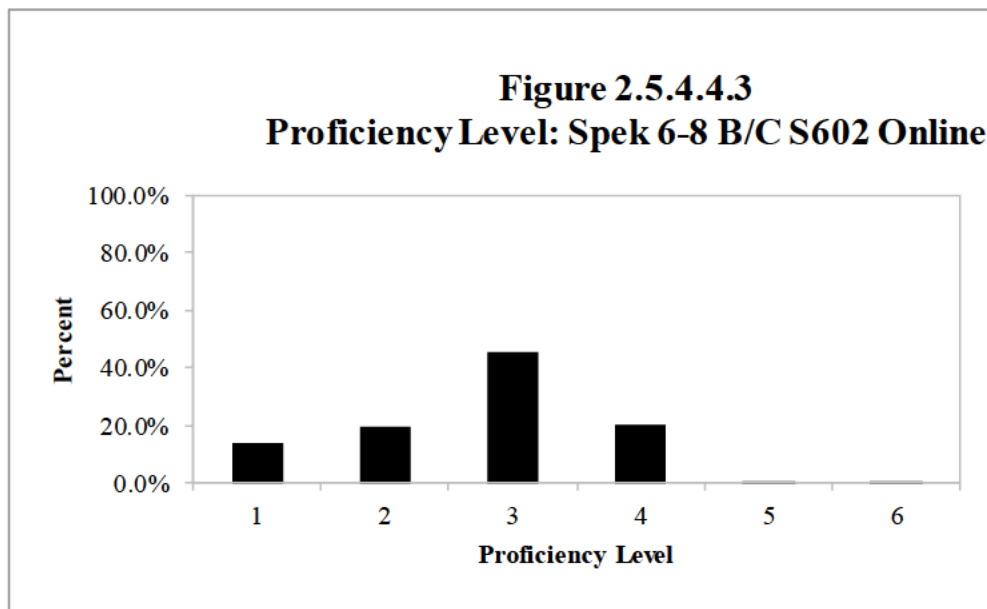


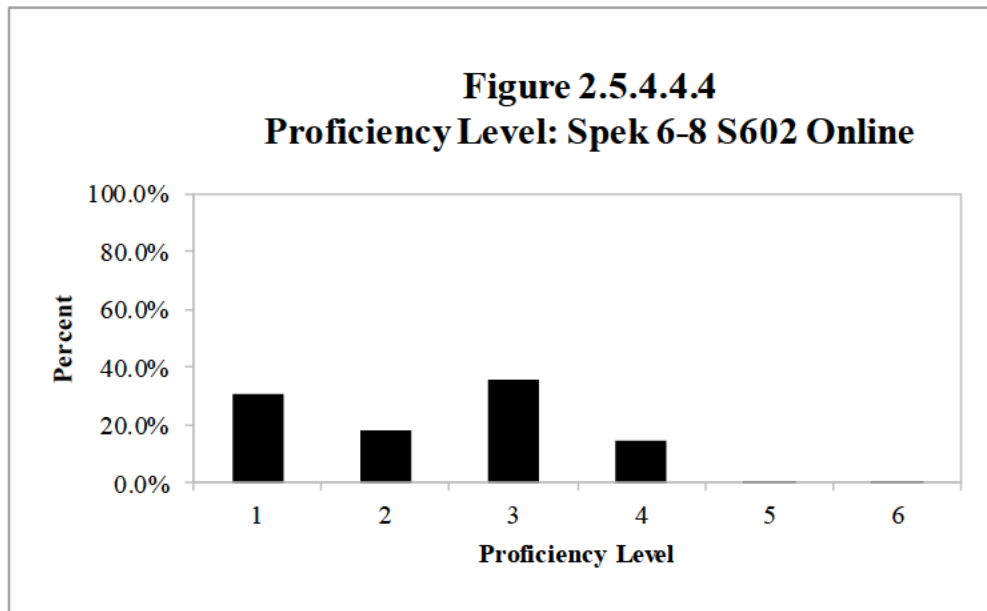
Table 2.5.4.4.4

Proficiency Level Distribution: Spek 6–8 S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	40,496	27.78%	49,536	32.74%	48,411	32.06%	138,443	30.90%
2	29,992	20.58%	28,347	18.73%	23,211	15.37%	81,550	18.20%
3	56,023	38.43%	47,069	31.11%	57,048	37.78%	160,140	35.74%
4	18,680	12.81%	25,166	16.63%	21,627	14.32%	65,473	14.61%
5	541	0.37%	1,111	0.73%	505	0.33%	2,157	0.48%
6	35	0.02%	93	0.06%	186	0.12%	314	0.07%
Total	145,767	100.00%	151,322	100.00%	150,988	100.00%	448,077	100.00%

Figure 2.5.4.4.4

Proficiency Level: Spek 6–8 S602 Online



2.5.4.5 Grades 9–12

Table 2.5.4.5.1

Proficiency Level Distribution: Spek 9–12 Pre-A S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	9,161	100.00%	9,342	100.00%	11,193	100.00%	8,498	100.00%	38,194	100.00%
Total	9,161	100.00%	9,342	100.00%	11,193	100.00%	8,498	100.00%	38,194	100.00%

Figure 2.5.4.5.1

Proficiency Level: Spek 9–12 Pre-A S602 Online

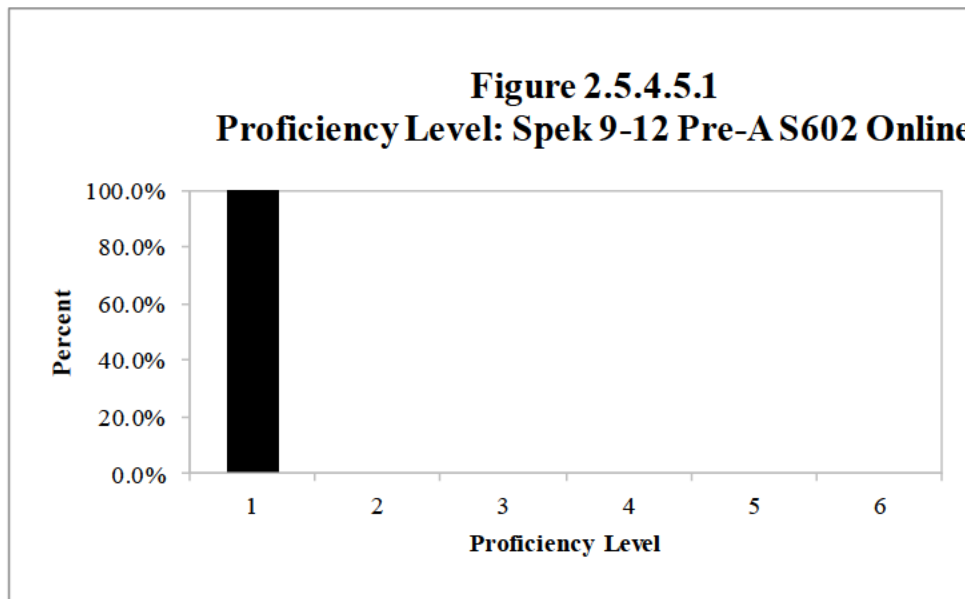


Table 2.5.4.5.2

Proficiency Level Distribution: Spek 9–12 A S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	54,711	57.97%	34,613	57.39%	15,116	60.62%	15,044	38.17%	119,484	54.55%
2	16,813	17.82%	11,061	18.34%	4,741	19.01%	16,268	41.28%	48,883	22.32%
3	21,905	23.21%	13,973	23.17%	4,871	19.53%	7,842	19.90%	48,591	22.18%
4	917	0.97%	661	1.10%	207	0.83%	257	0.65%	2,042	0.93%
5	27	0.03%	0	0.00%	0	0.00%	0	0.00%	27	0.01%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	94,373	100.00%	60,308	100.00%	24,935	100.00%	39,411	100.00%	219,027	100.00%

Figure 2.5.4.5.2

Proficiency Level: Spek 9–12 A S602 Online

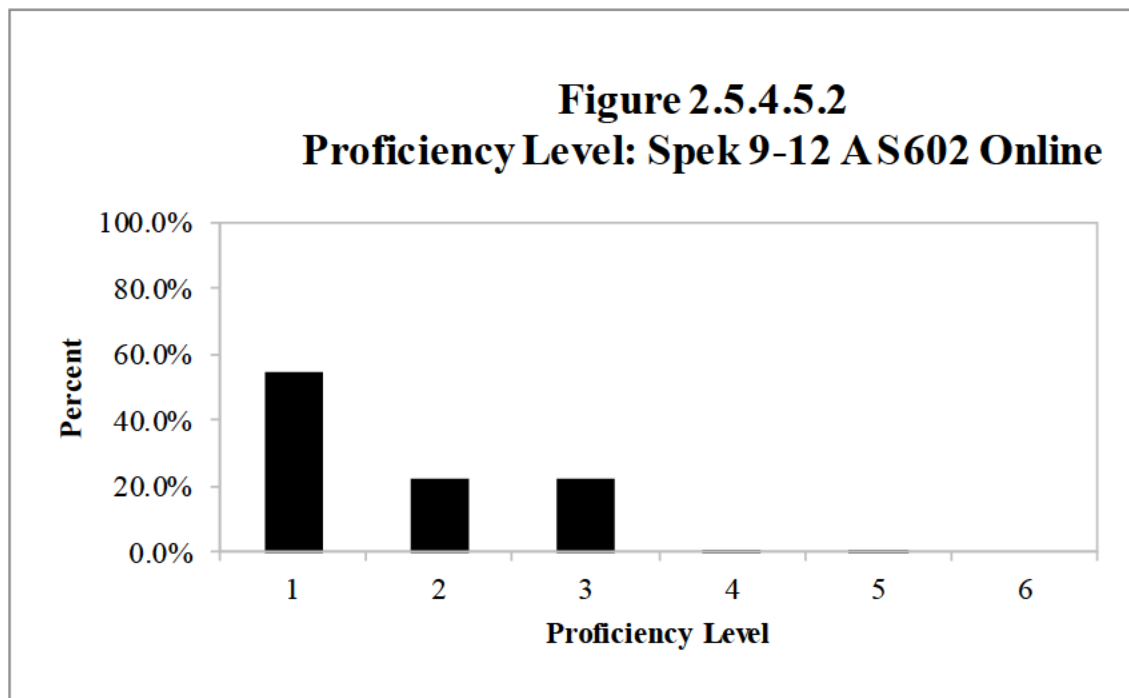


Table 2.5.4.5.3

Proficiency Level Distribution: Spek 9–12 B/C S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	9,430	15.23%	11,251	15.69%	14,682	19.34%	5,632	16.38%	40,995	16.81%
2	16,004	25.84%	17,211	24.01%	17,853	23.51%	9,089	26.43%	60,157	24.66%
3	31,713	51.21%	39,808	55.53%	39,560	52.10%	18,469	53.70%	129,550	53.11%
4	4,630	7.48%	3,205	4.47%	3,551	4.68%	1,032	3.00%	12,418	5.09%
5	106	0.17%	165	0.23%	220	0.29%	121	0.35%	612	0.25%
6	45	0.07%	52	0.07%	66	0.09%	47	0.14%	210	0.09%
Total	61,928	100.00%	71,692	100.00%	75,932	100.00%	34,390	100.00%	243,942	100.00%

Figure 2.5.4.5.3

Proficiency Level: Spek 9–12 B/C S602 Online

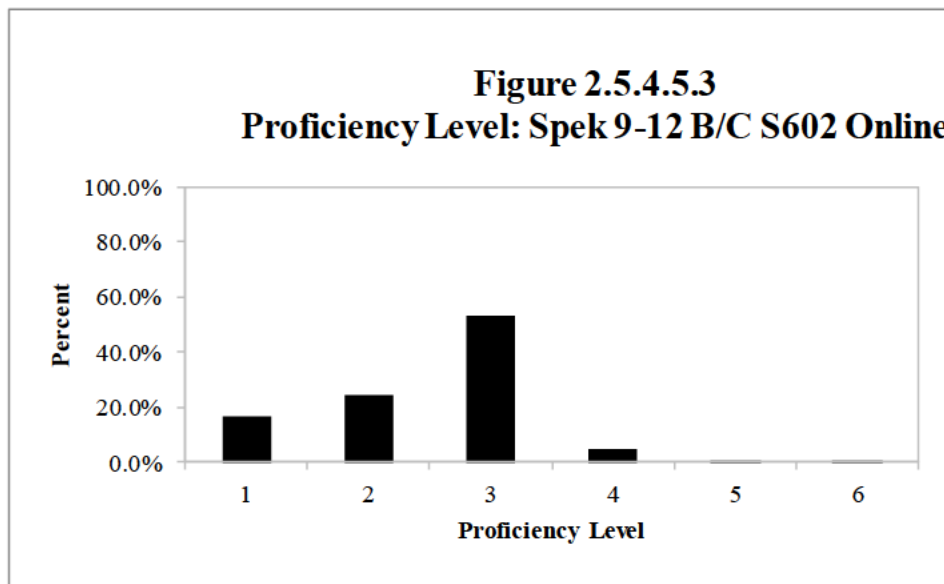


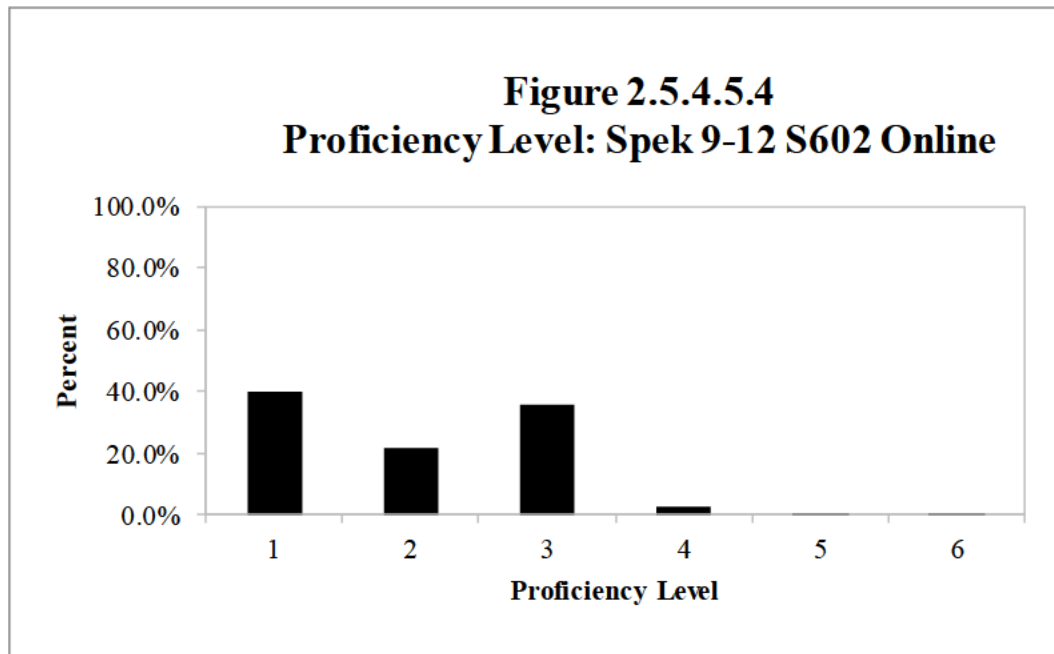
Table 2.5.4.5.4

Proficiency Level Distribution: Spek 9–12 S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	73,302	44.30%	55,206	39.06%	40,991	36.58%	29,174	35.45%	198,673	39.64%
2	32,817	19.83%	28,272	20.00%	22,594	20.16%	25,357	30.81%	109,040	21.76%
3	53,618	32.41%	53,781	38.05%	44,431	39.65%	26,311	31.97%	178,141	35.55%
4	5,547	3.35%	3,866	2.74%	3,758	3.35%	1,289	1.57%	14,460	2.89%
5	133	0.08%	165	0.12%	220	0.20%	121	0.15%	639	0.13%
6	45	0.03%	52	0.04%	66	0.06%	47	0.06%	210	0.04%
Total	165,462	100.00%	141,342	100.00%	112,060	100.00%	82,299	100.00%	501,163	100.00%

Figure 2.5.4.5.4

Proficiency Level: Spek 9–12 S602 Online



2.6 *Raw Score to Scale Score to Proficiency Level Conversion for Speaking and Writing*

This section presents raw score to scale score conversions and associated proficiency levels for the test forms for Speaking and Writing.

The first column in the tables shows all possible raw scores. The second column shows the corresponding scale score. The third column shows the conditional standard error of measurement (CSEM) in the metric of the scale score, multiplied by 1.96. The resulting number (CSEM x 1.96) is used to construct the confidence band as reported on students' score reports. For example, if a student receives a scale score of 199 and if the CSEM multiplied by 1.96 is 45, then there is a 95% chance that the student's true scale score will be found somewhere between 154 and 244. For additional detail on conditional standard error of measurement, see Section 5, Reliability. Following the CSEM, columns provide the proficiency level interpretation for each grade in the grade-level cluster.

Performances that gain very few score points, and performances from students who gain all or almost all the score points, will have high CSEM values. The model does not precisely estimate these students' abilities; they may be well below or well above the range that is measured by the test and therefore the error of measurement is large. We provide further detail on the CSEM as it relates to the interpretation of student performances in Section 5.3, which provides CSEM values for proficiency level cuts.

Note that we truncate raw scores of zero where necessary so that the lowest scale score given is the scale score corresponding to a proficiency level score of 1.0.

2.6.1 Listening

The ACCESS Online Listening test is a multistage adaptive assessment. As students do not all take the same set of items in the test, raw to scale score conversion tables are not presented.

2.6.2 Reading

The ACCESS Online Reading test is a multistage adaptive assessment. As students do not all take the same set of items in the test, raw to scale score conversion tables are not presented.

2.6.3 Writing

2.6.3.1 Grade 1

Table 2.6.3.1.1

Raw Score to Scale Score to Proficiency Level Conversion: Writ 1 A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	111	256	1.0
1	191	45	1.6
2	205	33	1.7
3	214	29	1.8
4	222	28	1.8
5	230	29	1.9
6	239	31	2.0
7	250	34	2.3
8	263	38	2.6
9	279	41	3.0
10	296	42	3.3
11	314	42	3.6
12	331	40	3.9
13	346	38	4.2
14	360	36	4.5
15	373	37	4.8
16	387	40	5.2
17	407	52	6.0
18	439	94	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.3.1.2**Raw Score to Scale Score to Proficiency Level Conversion: Writ 1 B/C S602 Online**

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	111	256	1.0
1	209	45	1.7
2	223	33	1.8
3	232	29	1.9
4	240	28	2.0
5	248	29	2.2
6	257	31	2.5
7	267	34	2.7
8	281	38	3.0
9	297	41	3.3
10	314	42	3.6
11	332	42	3.9
12	349	40	4.2
13	364	38	4.6
14	377	36	4.8
15	391	37	5.3
16	405	40	6.0
17	425	52	6.0
18	457	94	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

2.6.3.2 Grades 2–3

Table 2.6.3.2.1

Raw Score to Scale Score to Proficiency Level Conversion: Writ 2–3 A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G2	PL for G3
0	133	256	1.0	1.0
1	206	45	1.7	1.6
2	220	32	1.8	1.7
3	229	28	1.8	1.8
4	236	27	1.9	1.9
5	244	28	2.0	1.9
6	253	31	2.2	2.1
7	263	35	2.5	2.4
8	277	39	2.9	2.8
9	293	41	3.2	3.1
10	310	42	3.5	3.4
11	328	42	3.7	3.7
12	345	40	4.0	3.9
13	360	38	4.4	4.2
14	374	36	4.7	4.5
15	387	36	4.9	4.8
16	401	40	5.5	5.2
17	421	52	6.0	6.0
18	452	94	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.3.2.2**Raw Score to Scale Score to Proficiency Level Conversion: Writ 2–3 B/C S602 Online**

Raw Score	Scale Score	CSEM x 1.96	PL for G2	PL for G3
0	133	256	1.0	1.0
1	213	45	1.7	1.7
2	227	32	1.8	1.8
3	236	28	1.9	1.9
4	243	27	2.0	1.9
5	251	28	2.2	2.1
6	259	31	2.4	2.3
7	270	35	2.7	2.6
8	284	39	3.0	3.0
9	300	41	3.3	3.2
10	317	42	3.6	3.5
11	335	42	3.9	3.8
12	352	40	4.2	4.1
13	367	38	4.5	4.4
14	381	36	4.8	4.7
15	394	36	5.2	5.0
16	408	40	5.8	5.5
17	427	52	6.0	6.0
18	459	94	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

2.6.3.3 Grades 4–5

Table 2.6.3.3.1

Raw Score to Scale Score to Proficiency Level Conversion: Writ 4–5 A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	155	256	1.0	1.0
1	236	45	1.7	1.7
2	250	32	1.8	1.8
3	259	28	1.9	1.9
4	267	27	2.0	2.0
5	274	28	2.3	2.2
6	283	31	2.7	2.6
7	293	35	3.0	3.0
8	307	39	3.3	3.2
9	323	41	3.5	3.4
10	340	42	3.8	3.7
11	358	42	4.1	4.0
12	375	40	4.4	4.3
13	390	38	4.7	4.6
14	404	36	5.1	4.9
15	417	36	5.6	5.3
16	431	40	6.0	5.9
17	451	52	6.0	6.0
18	482	94	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.3.3.2**Raw Score to Scale Score to Proficiency Level Conversion: Writ 4–5 B/C S602 Online**

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	155	256	1.0	1.0
1	260	45	1.9	1.9
2	274	32	2.3	2.2
3	283	28	2.7	2.6
4	290	27	3.0	2.8
5	298	28	3.1	3.0
6	307	31	3.3	3.2
7	317	35	3.4	3.3
8	331	39	3.6	3.6
9	347	41	3.9	3.8
10	364	42	4.2	4.1
11	382	42	4.6	4.5
12	399	40	4.9	4.8
13	414	38	5.5	5.2
14	428	36	6.0	5.8
15	441	36	6.0	6.0
16	455	40	6.0	6.0
17	475	52	6.0	6.0
18	506	94	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

2.6.3.4 Grades 6–8

Table 2.6.3.4.1

Raw Score to Scale Score to Proficiency Level Conversion: Writ 6–8 A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	188	126	1.2	1.1	1.0
1	228	45	1.6	1.5	1.4
2	242	32	1.7	1.6	1.5
3	251	28	1.8	1.7	1.6
4	259	27	1.9	1.8	1.7
5	267	28	1.9	1.9	1.8
6	275	31	2.2	2.0	1.9
7	286	35	2.6	2.4	2.1
8	299	39	3.0	2.8	2.6
9	315	41	3.2	3.1	3.0
10	333	42	3.5	3.4	3.3
11	351	42	3.8	3.7	3.6
12	368	40	4.1	4.0	3.9
13	383	38	4.4	4.3	4.2
14	396	36	4.6	4.5	4.4
15	409	36	4.9	4.8	4.7
16	423	40	5.3	5.1	4.9
17	443	52	6.0	5.7	5.5
18	475	94	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.3.4.2**Raw Score to Scale Score to Proficiency Level Conversion: Writ 6–8 B/C S602 Online**

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	188	183	1.2	1.1	1.0
1	245	45	1.7	1.7	1.6
2	259	32	1.9	1.8	1.7
3	268	28	2.0	1.9	1.8
4	276	27	2.2	2.0	1.9
5	284	28	2.5	2.3	2.1
6	292	31	2.8	2.5	2.3
7	303	35	3.0	2.9	2.7
8	316	39	3.2	3.1	3.0
9	332	41	3.5	3.4	3.3
10	350	42	3.8	3.7	3.6
11	368	42	4.1	4.0	3.9
12	385	40	4.4	4.3	4.2
13	400	38	4.7	4.6	4.5
14	413	36	5.0	4.8	4.7
15	426	36	5.4	5.2	5.0
16	440	40	5.9	5.6	5.4
17	460	52	6.0	6.0	6.0
18	492	94	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

2.6.3.5 Grades 9–12

Table 2.6.3.5.1

Raw Score to Scale Score to Proficiency Level Conversion: Writ 9–12 A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	232	92	1.3	1.2	1.1	1.0
1	259	45	1.6	1.5	1.4	1.3
2	273	32	1.8	1.7	1.5	1.4
3	282	29	1.9	1.8	1.7	1.5
4	290	28	2.0	1.9	1.7	1.6
5	298	28	2.3	2.0	1.8	1.7
6	306	31	2.5	2.2	1.9	1.8
7	317	35	2.9	2.6	2.3	1.9
8	330	39	3.1	3.0	2.8	2.4
9	346	41	3.4	3.3	3.1	3.0
10	364	42	3.7	3.6	3.5	3.3
11	382	42	4.0	3.9	3.8	3.7
12	398	40	4.3	4.2	4.1	4.0
13	414	38	4.6	4.5	4.4	4.3
14	427	36	4.9	4.8	4.7	4.5
15	440	36	5.2	5.0	4.9	4.8
16	454	40	5.6	5.4	5.2	5.1
17	474	52	6.0	5.8	5.6	5.5
18	506	94	6.0	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.3.5.2**Raw Score to Scale Score to Proficiency Level Conversion: Writ 9–12 B/C S602 Online**

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	232	87	1.3	1.2	1.1	1.0
1	257	45	1.6	1.5	1.4	1.2
2	271	33	1.8	1.6	1.5	1.4
3	281	30	1.9	1.8	1.6	1.5
4	289	28	2.0	1.8	1.7	1.6
5	298	29	2.3	2.0	1.8	1.7
6	307	31	2.6	2.3	1.9	1.8
7	317	34	2.9	2.6	2.3	1.9
8	330	38	3.1	3.0	2.8	2.4
9	346	41	3.4	3.3	3.1	3.0
10	363	42	3.7	3.6	3.5	3.3
11	381	42	4.0	3.9	3.8	3.6
12	398	40	4.3	4.2	4.1	4.0
13	413	38	4.6	4.5	4.4	4.3
14	427	37	4.9	4.8	4.7	4.5
15	440	37	5.2	5.0	4.9	4.8
16	455	40	5.6	5.4	5.2	5.1
17	475	52	6.0	5.9	5.6	5.5
18	506	94	6.0	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

2.6.4 Speaking

2.6.4.1 Grade 1

Table 2.6.4.1.1

Raw Score to Scale Score to Proficiency Level Conversion: Spek 1 Pre-A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	106	44	1.0
1	106	44	1.0
2	115	40	1.0
3	128	37	1.2
4	141	40	1.3
5	154	48	1.4
6	167	61	1.6

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.4.1.2

Raw Score to Scale Score to Proficiency Level Conversion: Spek 1 A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	106	45	1.0
1	106	45	1.0
2	115	39	1.0
3	127	34	1.2
4	137	32	1.3
5	146	31	1.4
6	155	32	1.4
7	165	33	1.5
8	175	35	1.6
9	187	37	1.8
10	201	41	1.9
11	218	48	2.2
12	242	54	2.6
13	269	52	3.1
14	291	47	3.6
15	310	46	3.9
16	331	50	4.4
17	352	59	4.8
18	373	75	5.2

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.4.1.3**Raw Score to Scale Score to Proficiency Level Conversion: Spek 1 B/C S602 Online**

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	106	77	1.0
1	124	58	1.1
2	145	42	1.3
3	159	36	1.5
4	170	33	1.6
5	180	31	1.7
6	188	31	1.8
7	197	31	1.9
8	206	31	2.0
9	215	33	2.1
10	226	35	2.3
11	237	37	2.5
12	250	39	2.8
13	264	38	3.0
14	277	37	3.3
15	289	35	3.5
16	300	34	3.7
17	310	33	3.9
18	320	33	4.1
19	330	34	4.3
20	341	35	4.6
21	353	39	4.8
22	365	43	5.0
23	377	49	5.3
24	403	69	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

2.6.4.2 Grades 2–3

Table 2.6.4.2.1

Raw Score to Scale Score to Proficiency Level Conversion: Spek 2–3 Pre-A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G2	PL for G3
0	118	38	1.0	1.0
1	118	38	1.0	1.0
2	118	38	1.0	1.0
3	125	37	1.1	1.0
4	138	40	1.2	1.1
5	151	47	1.3	1.2
6	164	60	1.4	1.3

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.4.2.2

Raw Score to Scale Score to Proficiency Level Conversion: Spek 2–3 A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G2	PL for G3
0	118	39	1.0	1.0
1	118	39	1.0	1.0
2	118	39	1.0	1.0
3	130	34	1.1	1.1
4	141	33	1.2	1.1
5	151	34	1.3	1.2
6	162	35	1.4	1.3
7	174	37	1.5	1.4
8	187	38	1.6	1.5
9	201	40	1.8	1.7
10	216	43	1.9	1.8
11	235	48	2.2	2.0
12	259	54	2.7	2.5
13	285	52	3.2	3.0
14	308	48	3.7	3.5
15	328	47	4.1	3.9
16	349	50	4.5	4.3
17	370	59	4.9	4.7
18	391	75	5.4	5.1

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.4.2.3**Raw Score to Scale Score to Proficiency Level Conversion: Spek 2–3 B/C S602 Online**

Raw Score	Scale Score	CSEM x 1.96	PL for G2	PL for G3
0	118	84	1.0	1.0
1	141	58	1.2	1.1
2	163	43	1.4	1.3
3	177	37	1.6	1.5
4	188	34	1.7	1.6
5	198	32	1.7	1.6
6	207	31	1.8	1.7
7	216	31	1.9	1.8
8	225	32	2.0	1.9
9	235	33	2.2	2.0
10	245	35	2.4	2.2
11	257	37	2.6	2.4
12	270	38	2.9	2.7
13	283	38	3.2	3.0
14	296	37	3.4	3.2
15	307	35	3.6	3.4
16	318	34	3.9	3.7
17	329	34	4.1	3.9
18	339	34	4.3	4.1
19	350	34	4.5	4.3
20	361	36	4.7	4.5
21	374	39	5.0	4.7
22	387	44	5.3	5.0
23	400	51	5.6	5.3
24	425	71	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

2.6.4.3 Grades 4–5

Table 2.6.4.3.1

Raw Score to Scale Score to Proficiency Level Conversion: Spek 4–5 Pre-A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	130	44	1.0	1.0
1	130	44	1.0	1.0
2	137	40	1.1	1.0
3	151	38	1.2	1.1
4	164	40	1.3	1.2
5	177	48	1.4	1.3
6	190	60	1.5	1.4

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.4.3.2

Raw Score to Scale Score to Proficiency Level Conversion: Spek 4–5 A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	130	49	1.0	1.0
1	130	49	1.0	1.0
2	144	40	1.1	1.1
3	158	36	1.2	1.2
4	169	36	1.3	1.3
5	181	37	1.4	1.3
6	194	39	1.5	1.5
7	209	40	1.6	1.6
8	224	41	1.8	1.7
9	239	41	1.9	1.8
10	255	44	2.1	1.9
11	275	49	2.6	2.3
12	299	54	3.1	2.9
13	325	52	3.6	3.4
14	348	48	4.1	3.9
15	368	47	4.4	4.3
16	389	50	4.8	4.6
17	410	59	5.3	5.0
18	431	75	5.8	5.6

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.4.3.3**Raw Score to Scale Score to Proficiency Level Conversion: Spek 4–5 B/C S602 Online**

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	130	144	1.0	1.0
1	185	59	1.5	1.4
2	208	43	1.6	1.6
3	222	37	1.8	1.7
4	234	34	1.9	1.8
5	244	33	1.9	1.8
6	253	32	2.1	1.9
7	263	32	2.3	2.1
8	272	32	2.5	2.3
9	281	33	2.7	2.5
10	292	34	2.9	2.7
11	303	36	3.2	3.0
12	316	37	3.4	3.2
13	329	38	3.7	3.5
14	341	37	3.9	3.8
15	353	35	4.2	4.0
16	364	34	4.4	4.2
17	375	34	4.6	4.4
18	385	34	4.7	4.6
19	396	35	4.9	4.8
20	407	36	5.2	5.0
21	420	39	5.6	5.3
22	433	44	5.9	5.7
23	446	51	6.0	6.0
24	459	60	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

2.6.4.4 Grades 6–8

Table 2.6.4.4.1

Raw Score to Scale Score to Proficiency Level Conversion: Spek 6–8 Pre-A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	148	47	1.0	1.0	1.0
1	148	47	1.0	1.0	1.0
2	160	40	1.1	1.1	1.0
3	173	37	1.2	1.2	1.1
4	186	40	1.3	1.3	1.2
5	199	48	1.4	1.4	1.3
6	212	60	1.5	1.5	1.4

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.4.4.2

Raw Score to Scale Score to Proficiency Level Conversion: Spek 6–8 A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	148	45	1.0	1.0	1.0
1	148	45	1.0	1.0	1.0
2	158	39	1.1	1.1	1.0
3	170	34	1.2	1.2	1.1
4	180	33	1.3	1.2	1.2
5	190	34	1.4	1.3	1.3
6	202	36	1.5	1.4	1.3
7	214	38	1.5	1.5	1.4
8	228	39	1.6	1.6	1.5
9	241	40	1.7	1.7	1.6
10	257	43	1.9	1.8	1.8
11	275	49	2.1	1.9	1.9
12	300	55	2.7	2.5	2.4
13	327	52	3.3	3.1	3.0
14	349	47	3.7	3.6	3.4
15	369	46	4.1	4.0	3.8
16	390	50	4.5	4.3	4.2
17	411	59	4.8	4.7	4.6
18	432	76	5.4	5.2	4.9

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.4.4.3**Raw Score to Scale Score to Proficiency Level Conversion: Spek 6–8 B/C S602 Online**

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	148	105	1.0	1.0	1.0
1	184	58	1.3	1.3	1.2
2	205	42	1.5	1.4	1.4
3	219	36	1.6	1.5	1.5
4	230	33	1.7	1.6	1.6
5	239	31	1.7	1.7	1.6
6	248	31	1.8	1.7	1.7
7	257	31	1.9	1.8	1.8
8	265	31	1.9	1.9	1.8
9	275	33	2.1	1.9	1.9
10	285	35	2.4	2.2	2.0
11	297	37	2.6	2.5	2.3
12	310	39	3.0	2.8	2.6
13	324	38	3.2	3.1	3.0
14	337	37	3.5	3.3	3.2
15	349	35	3.7	3.6	3.4
16	359	34	3.9	3.8	3.6
17	370	33	4.1	4.0	3.8
18	380	33	4.3	4.1	4.0
19	390	34	4.5	4.3	4.2
20	401	35	4.7	4.5	4.4
21	413	39	4.9	4.7	4.6
22	425	43	5.2	5.0	4.8
23	437	49	5.5	5.3	5.1
24	463	70	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

2.6.4.5 Grades 9–12

Table 2.6.4.5.1

Raw Score to Scale Score to Proficiency Level Conversion: Spek 9–12 Pre-A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	172	38	1.1	1.0	1.0	1.0
1	172	38	1.1	1.0	1.0	1.0
2	172	38	1.1	1.0	1.0	1.0
3	180	37	1.1	1.1	1.1	1.0
4	193	40	1.2	1.2	1.2	1.1
5	206	48	1.3	1.3	1.3	1.2
6	219	61	1.4	1.4	1.3	1.3

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.4.5.2

Raw Score to Scale Score to Proficiency Level Conversion: Spek 9–12 A S602 Online

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	172	36	1.1	1.0	1.0	1.0
1	172	36	1.1	1.0	1.0	1.0
2	172	36	1.1	1.0	1.0	1.0
3	176	35	1.1	1.1	1.0	1.0
4	186	34	1.2	1.1	1.1	1.1
5	197	35	1.3	1.2	1.2	1.1
6	209	37	1.4	1.3	1.3	1.2
7	223	39	1.5	1.4	1.4	1.3
8	237	40	1.6	1.5	1.5	1.5
9	252	41	1.7	1.6	1.6	1.6
10	268	43	1.8	1.8	1.7	1.7
11	287	48	1.9	1.9	1.9	1.8
12	311	54	2.5	2.4	2.3	2.2
13	338	52	3.1	3.0	3.0	2.9
14	360	48	3.5	3.4	3.3	3.3
15	381	47	3.9	3.8	3.6	3.6
16	402	50	4.3	4.1	4.0	3.9
17	423	60	4.6	4.5	4.4	4.3
18	444	75	5.1	4.9	4.8	4.7

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.4.5.3**Raw Score to Scale Score to Proficiency Level Conversion: Spek 9–12 B/C S602 Online**

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	172	72	1.1	1.0	1.0	1.0
1	185	58	1.2	1.1	1.1	1.1
2	207	43	1.3	1.3	1.3	1.2
3	221	37	1.4	1.4	1.4	1.3
4	232	33	1.5	1.5	1.4	1.4
5	242	32	1.6	1.6	1.5	1.5
6	251	31	1.7	1.6	1.6	1.6
7	259	31	1.7	1.7	1.6	1.6
8	268	31	1.8	1.8	1.7	1.7
9	278	33	1.9	1.8	1.8	1.8
10	288	35	1.9	1.9	1.9	1.8
11	300	37	2.2	2.1	2.0	1.9
12	313	38	2.6	2.4	2.3	2.2
13	326	38	2.9	2.8	2.7	2.6
14	339	37	3.1	3.1	3.0	2.9
15	351	35	3.4	3.3	3.2	3.1
16	362	34	3.5	3.4	3.3	3.3
17	372	33	3.7	3.6	3.5	3.4
18	382	33	3.9	3.8	3.7	3.6
19	392	34	4.1	3.9	3.8	3.7
20	403	36	4.3	4.1	4.0	3.9
21	416	39	4.5	4.4	4.3	4.2
22	429	44	4.8	4.6	4.5	4.4
23	455	60	5.5	5.3	5.1	5.0
24	476	81	6.0	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

2.7 Equating Summary

Each year a certain number of items and tasks on the ACCESS for ELLs Online test form are new, as determined by the refreshment plan for that series. For Series 602, we refreshed all four domains.

For the Listening and Reading domains, WIDA implements a multiyear targeted refreshment plan to optimize the multistage computerized adaptive item pools and to ensure that we do not use these folders in the pools too long, thus overexposing them. In the spring of 2021, WIDA and CAL assessment experts reviewed the 601 Listening and Reading item pools and identified folders that they believed the team should refresh for Series 602, according to the targeted

refreshment plan. To meet these Series 602 targets, DRC field tested 87 Listening folders and 111 Reading folders.

For the Writing and Speaking domains, which are shorter, performance based, and which have additional content and exposure considerations in terms of task refreshment, WIDA and CAL assessment experts created the refreshment plan three years earlier to ensure that the test development effort could accommodate the refreshment target set for each series.

The Writing test consists of two sets of operational tasks that target four of the five WIDA ELD Standards. The first set targets Standard 2: Language of Language Arts and Standard 5: Language of Social Studies. The second set targets Standard 3: Language of Mathematics and Standard 4: Language of Science. The test creators designed each set of operational tasks, as well as each set of anchor tasks, to measure student performance across the entire proficiency scale, from PL 1 to PL 6. We refresh one of the two sets each year, on an alternating schedule, so the two WIDA ELD Standards that the anchor tasks target alternate from year to year.

The Speaking test consists of three sets of operational tasks that target all five WIDA ELD Standards. The first set targets Standard 1: Social and Instructional Language. The second set targets Standard 2: Language of Language Arts and Standard 5: Language of Social Studies. The third set targets Standard 3: Language of Mathematics and Standard 4: Language of Science. The test creators designed each set of operational tasks, as well as each set of anchor tasks, to measure student performance across the entire proficiency scale, from PL 1 to PL 6. Generally, we refresh one (or two) of the three sets each year on a rotating schedule, so the two WIDA ELD Standards that the anchor tasks target also rotate from year to year. This allows for the Speaking test to be of manageable length and still contain embedded field test tasks, in consideration of the seat time required of students to complete each Speaking performance task. We refreshed two panels, or six tasks, for Series 602.

When we consider the sets of anchor tasks for the Speaking and Writing tests, it is important to note the overall assessment construct when we further consider the distribution of anchor tasks. The overarching goal of ACCESS for ELLs Online is to measure academic English language proficiency of students in each of the four domains. WIDA measures English language proficiency using a 6-level scale, which is defined in the WIDA Performance Definitions for the receptive domains (Listening and Reading) and productive domains (Speaking and Writing). WIDA does not have performance definitions that define a proficiency scale for each of the WIDA Standard Statements (e.g., no performance definitions exist specifically for Social and Instructional Language or the Language of Math). Given that proficiency in the WIDA Standard Statements is not defined, ACCESS for ELLs does not measure proficiency in the WIDA Standard Statements, and thus WIDA does not report proficiency scores for students at the level of the WIDA Standard Statements (see Part 1, Section 1.2). Therefore, it is not necessary for the anchor sets in Speaking and Writing to contain tasks that target all five of the WIDA Standard Statements. Rather, it is more important to ensure that each anchor task assesses the targeted proficiency levels so we can sufficiently claim that ACCESS for ELLs Online truly measures across the breadth of the proficiency scale.

We used an equating procedure, known as common item equating, to equate the results from the new item/task pool and forms to the older item/task pool and forms using the common

items/tasks, which are items/tasks that appear in both Series 601 and 602 for all domains. The characteristics of the common items/tasks were kept the same between series, as were the wording, formatting, and other test characteristics such as graphics. Furthermore, common items/tasks appeared in the same item/task sequence position as they appeared in the previous test series. In this procedure, we kept constant across both pools and test forms the difficulty measures for the items and tasks included on both the new and the old forms. In this way, the test user may employ the same frame of reference when interpreting students' scores on the newer test forms.

For the Listening and Reading domains, we used a pre-equating design to conduct the annual equating using student data collected from the Series 602 embedded field test (See Part 1, Section 2.3.2). This design allowed for Listening and Reading item parameters to be available for setting up the computer adaptive engine prior to operational administration. We included in the final analyses all the student data that was available at the time that we conducted these equating analyses. All common items between Series 602 and 601, except for four Reading items, are used as anchors and were maintained in that role if they met two criteria: (1) the item/task displayed adequate fit (i.e., item/task mean square infit and outfit measures were between -1.30 and 1.30, and (2) the item/task exhibited no C-level or CC-level DIF. Using these criteria, we did not need to remove any common items/tasks from the anchor sets for any of the Series 602 tests before conducting the equating analysis. Because we included all Series 601 operational items in the anchor set when conducting the annual equating, the content representation of the anchor set was not a concern. The four Reading items were dropped during 602 item selection meeting due to concerns of exposure issue, and hence 4 other folders were swapped into the 602 OP pool based on the decision made afterwards.

For both the Writing and Speaking tests, DRC implemented an embedded field test design (See Part 1, Section 2.3.2).

For the annual equating of the Writing test, DRC drew random samples of students from among those who had already taken the Writing test at the time of the draw, according to WIDA's predetermined sampling plan. When implementing that sampling plan, DRC drew a fixed number of students by grade-level cluster and tiered forms, where the number of students drawn was proportional to the population means of the number of students across previous series for the grade-level cluster and tiered forms.

For the annual equating of the Speaking test, DRC drew random samples of students from among those who had already taken the Speaking test at the time of the draw. When implementing that sampling plan, DRC drew a fixed number of students by grade-level cluster and tiered forms, where the number of students drawn was proportional to the population means of the number of students across previous series for the grade-level clusters. We included in the final analysis all the student data that was available at the time when we conducted our annual equating analyses.

The standard equating procedure involves anchoring all items/tasks common to Series 602 item/task pools and forms to their Series 601 values in the equating run, while the items and tasks parameters for new items and tasks were estimated. This procedure places the parameters of the new 602 items and tasks on the same scale as those of the 601 items and

tasks. For the Listening, Reading, and Speaking domains, we examined the displacement statistics of the anchored item/task after the first equating run. If the displacement statistics for any items and tasks is greater than the pre-established thresholds set by WIDA described below, the anchored items or tasks parameters will be re-estimated until the displacement statistics for all anchored items and tasks are below the thresholds. The **displacement statistic** shows the difference between the difficulty value of the anchored item/task and what its difficulty value would have been had we not anchored it. Smaller displacement statistics indicate more consistency between the item's (or task's) difficulty value on the Series 602 test form and on the Series 601 test form. Typically, displacements of less than 0.5 logits are unlikely to have much impact on measurement in a test instrument (Linacre, n.d.). For Listening and Reading items and P3 and P5 Speaking tasks, if this value was large (i.e., above 0.30 or below -0.30), that item was unanchored in the final equating run (i.e., it was treated as if it were a new item). For the Speaking P1 tasks, we used a slightly different displacement criterion (above 0.50 or below -0.50) since anchored P1 tasks from the Speaking domain have been found to be less stable than items and tasks from the other domains. Specifically, the test creators designed the Speaking P1 tasks to be very easy and therefore we can expect most students (98% to 99%) to get the full two points. As a result, the item difficulties for these P1 tasks are susceptible to small sampling fluctuations. A slight change in the percentages of students getting the full two points, due to sampling fluctuation, tends to cause the task difficulty values to change such that the displacement statistics will be out of the -0.3 and 0.3 range. If we were to use the same displacement criterion as other tasks, task difficulties for the P1 tasks would need to be re-estimated each time a slightly different sample is used to estimate them. Therefore, we used a more conservative estimate (-0.5 to 0.5) to evaluate the displacement statistics for the Speaking P1 tasks in order to ensure the stability of the Speaking scale scores. Since the Writing test has only one task anchored, there are no displacement statistics to evaluate.

Because of an item exposure issue of the Speaking equating sample, WIDA requested a modification to the equating procedure for the Speaking test. Specially, three new tasks (Task ID: 19928, 19935, and 19013) were exposed during the time the data of the equating sample were collected. Due to the concern that the equating sample's responses to these three tasks might have been compromised, CAL fixed the parameters of these tasks to their field test values instead of estimating them using the equating sample. For the rest of the anchored tasks, CAL evaluated their displacement statistics using normal procedure.

The tables that follow present a summary of the equating results. The first section of each table compares the current test (i.e., the Series 602 version of that item/task pool and test form) to the previous year's test (i.e., the Series 601 version of that item/task pool and test form). The table shows the number of items/tasks, the average item/task difficulty, the standard deviation of the item/task difficulty values, and the difficulty value of the easiest and hardest item/task on each test form. These values are in log-odd units, or **logits** (i.e., analyses carried out using Rasch measurement techniques, which produce equal-interval, linear measures expressed on a logit scale). In the domains of Listening and Reading, if the equating is successful, we would expect the average item difficulty values for the two series to be similar. This is true for these domains because they have many test items in the item pool, as well as large anchor sets.

Additionally, the Series 602 Writing domain tests consist of only two tasks, with only one task serving as an anchor between series. Therefore, we might expect some differences in the average difficulty values for the two Writing series. Similarly, we might expect some differences in the average difficulty values for the two Speaking series, as those test forms included only nine tasks, and one-third of the test served as the anchor between series.

The second section of each table presents information about the anchor items/tasks and shows the total number of possible anchors that we initially anchored to the values from the previous series, as well as the average item/task difficulty and the average standard deviation of the difficulty values for those items/tasks. Next, the table shows the number of items/tasks that we anchored in the final equating run, again with the average item/task difficulty and the average standard deviation of those difficulty values for those items/tasks. Finally, the table gives the percentage of items/tasks that served as anchors and their average displacement values. In general, the larger the number and the higher the percentage of items/tasks anchored and the closer their average displacement is to 0.00, the more trustworthy the equating results will be (Jones & Smith, 2006; Stahl & Muckle, 2007).

The third section of each table gives information about the anchor items/tasks, both by order of displacement statistics and by order of item/task difficulty. The displacement statistics provide information regarding the difference between the difficulty value of each anchored item/task and what that difficulty value would have been had we not anchored the item/task. Smaller displacement statistics indicate more consistency between the item's (or task's) difficulty value between the Series 602 test form and on the Series 601 test form. The anchor items/tasks appearing on a given test form should have a range of item/task difficulties that mirrors the range of item/task difficulties in the entire pool (Kolen & Brennan, 2004).

The tables for the Writing and Speaking domains have a fourth section, which provides the anchored **Rasch rating scale model step measures** for each task (also known as Rasch structure calibrations, step parameters, step calibrations, or Rasch-Andrich thresholds). Step measures identify the particular points along the student proficiency continuum where it is equally probable that a rater evaluating a student's response to a task would have assigned a score in either of two adjacent score categories. That is, a step measure indicates how likely it is for a student to receive a score in a particular score category relative to the adjacent score category on that scale. It is not a measure of the difficulty of the category (Linacre, 2004).

If the score categories are working as those who designed the scoring scale intended, the step measures should advance from step to step by at least 1.4 logits, but not more than 5.0 logits (Linacre, 2004). However, the required degree of advancement in the step measures lessens as the number of score categories increases. For practical purposes, advances of 1.4 logits are generally not required to be able to make valid inferences regarding a student's level of proficiency based on their score (Linacre, 2004).

If the step measures do not advance, then that indicates that the raters likely assigned few scores in one (or more) score categories, resulting in a set of "disordered" thresholds. When the frequency of scores that raters assigned in a category is low, then the step measure for that category will be imprecisely estimated and potentially unstable (Linacre, 2004).

For the Writing test forms, multiple tasks appeared on each form. We employed a rating scale model to analyze the scores that the raters assigned to students' written responses to those tasks. When using this model, we assumed that the raters similarly used the score categories when assigning scores to students' responses to both tasks included on the test form. That is, under this assumption, when Winsteps analyzed the students' Writing scores, it treated the 3s that raters assigned to students' responses to one task as equivalent to the 3s that raters assigned to students' responses on another task. Similarly, the computer program treats the 4s that raters assigned to students' responses to one task as equivalent to the 4s that raters assigned to students' responses on another task. Accordingly, the output from the Winsteps analysis reports a single set of step measures that applied to both the Writing tasks appearing on that test form. The Writing step measures advanced from step to step except from Step 1 to Step 2, which indicated that raters tended to assign fewer scores of 1 when compared with the other score categories. The advances in the step measures ranged from 0.17 logits (from Step 2 to Step 3) to 1.28 logits (from Step 6 to Step 7). While these findings do not signal optimal scoring scale functioning (i.e., the step measures did not advance from step to step by at least 1.4 logits), raters' use of the Writing Scoring Scale should still yield student scores that test users can meaningfully interpret (Linacre, 2004). To provide anchors for the calibration of new Writing tasks, to facilitate their placement onto the common WIDA score scale each year, we held the step measures constant.

For the Speaking test forms, we used a rating scale model to analyze the scores that raters assigned students' responses to all the PL 1 tasks, assuming that raters used the three score categories (0–2) on that scoring scale in a similar manner when evaluating students' oral responses to those tasks. Similarly, we used the same rating scale model to analyze the scores that raters assigned students' responses to the PL 3 and PL 5 tasks, assuming that raters used the five score categories (0–4) on that scoring scale in a similar manner when evaluating students' oral responses to those tasks. Therefore, the step measures for all PL 1 tasks were the same, and the step measures for all PL 3 and PL 5 tasks were the same. The Speaking step measures advanced from step to step for the PL 1 tasks and for the PL 3 and PL 5 tasks. For the PL 1 tasks, the step measures advanced by 1.12 logits from Step 1 to Step 2. For the PL 3 and PL 5 tasks, the advances in the step measures ranged from 0.85 logits (from Step 1 to Step 2) to 3.26 logits (from Step 2 to Step 3). While these findings do not signal optimal scoring scale functioning (i.e., the step measures did not all advance from step to step by at least 1.4 logits), raters' use of the two Speaking Scoring Scales should still yield student scores that test users can meaningfully interpret (Linacre, 2004). As with Writing, these constant step measures help to provide anchors in the calibration of new Speaking tasks, facilitating their placement onto the common WIDA score scale each year.

The tables in the next section of this report reveal that the average difficulty levels for the items appearing on the Series 602 Listening and Reading test forms were similar to those for the previous series for all grade-level clusters. For the Listening domain, the differences in the average difficulty levels ranged from -1.11 logits (for grade 1) to 1.87 logits (for grades 9–12). Similarly, for the Reading domain, the differences in the average difficulty levels ranged from -0.96 logits (for grade 1) to 2.37 logits (for grades 9–12). For each Listening and Reading test

form, the anchor items represented a wide range of difficulties that spanned nearly the entire item difficulty continuum.

The differences in the average difficulty levels for the tasks appearing on the Writing test forms for Series 602 and 601 were less than 0.20 logits for all grade-level clusters and tiers, except for grades 4–5 Tier B/C and grades 6–8 Tier B/C. For grades 4–5 Tier B/C, the difference was 0.21 logits and for grades 6–8 Tier B/C, the difference was 0.24.

The differences in the average difficulty levels for the tasks appearing on the Speaking test forms for Series 602 and 601 were less than 0.20 logits for all grade-level. For each Speaking test form, the anchor tasks represented a range of difficulties that spanned nearly the entire task difficulty continuum.

WIDA psychometricians reviewed the equating plans before CAL conducted the equating analyses. The WIDA psychometricians then reviewed the equating results at the conclusion of the equating project to ensure that the equating was carried out correctly and the results were deemed reasonable. Besides the evidence listed above to the success of the equating results, WIDA and CAL psychometricians compare scoring tables across years to ensure that scores are comparable across test series, which demonstrates that the tests are comparable across series. In addition, WIDA and CAL psychometricians reviewed the annual equating results and identified issues that they felt they needed to bring to the attention of the WIDA Technical Advisory Committee.

2.7.1 Listening

2.7.1.1 Grade 1

Table 2.7.1.1

Equating Summary: List 1 S602 Online

Table 2.7.1.1

Equating Summary: List 1 S602 Online

Comparison of Forms	Form 602		Form 601			
	No. of Items	Average Difficulty (Std. Dev.)	No. of Items	Average Difficulty (Std. Dev.)		
	54	-1.11 (1.07)	54	-1.11 (1.06)		
	Easiest	Hardest	Easiest	Hardest		
	-3.59	1.46	-3.59	0.96		
Anchoring Items	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	46	-1.08 (1.04)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	46	-1.08 (1.04)				
	Percentage Anchors	Average Displacement				
	85%	0.02				
Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	17814	-1.12	-0.29	20909	-3.59	-0.14
	14897	-1.30	-0.28	14952	-3.03	-0.01
	18843	-0.99	-0.24	18841	-3.01	0.10
	18891	0.20	-0.18	13889	-2.96	0.11
	16559	0.50	-0.17	13891	-2.55	0.09
	18889	-0.69	-0.15	17813	-2.32	0.03
	17788	0.01	-0.15	13890	-2.23	0.24
	16533	-0.47	-0.14	18842	-2.16	0.28
	20909	-3.59	-0.14	20168	-1.99	0.04
	17793	-0.27	-0.10	16531	-1.79	0.00
	16560	-0.02	-0.08	17815	-1.76	0.18
	16642	-0.74	-0.08	14951	-1.68	0.24
	18890	0.18	-0.07	13900	-1.63	-0.06
	16641	-0.86	-0.06	20167	-1.63	0.03
	13899	-1.15	-0.06	20292	-1.58	0.05

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	13900	-1.63	-0.06	17791	-1.57	0.11
	20166	-0.24	-0.03	14899	-1.50	0.03
	19330	0.21	-0.03	14953	-1.43	0.20
	16558	-0.15	-0.02	14897	-1.30	-0.28
	14952	-3.03	-0.01	16535	-1.22	0.09
	20291	-1.06	-0.01	13899	-1.15	-0.06
	16531	-1.79	0.00	17814	-1.12	-0.29
	20167	-1.63	0.03	20291	-1.06	-0.01
	14899	-1.50	0.03	18843	-0.99	-0.24
	17813	-2.32	0.03	13898	-0.94	0.10
	20168	-1.99	0.04	16641	-0.86	-0.06
	20292	-1.58	0.05	19514	-0.77	0.09
	16640	-0.25	0.06	16642	-0.74	-0.08
	19513	0.12	0.08	18889	-0.69	-0.15
	19514	-0.77	0.09	16533	-0.47	-0.14
	20293	-0.22	0.09	19332	-0.32	0.23
	13891	-2.55	0.09	20245	-0.31	0.25
	16535	-1.22	0.09	17793	-0.27	-0.10
	18841	-3.01	0.10	16640	-0.25	0.06
	13898	-0.94	0.10	20166	-0.24	-0.03
	13889	-2.96	0.11	20293	-0.22	0.09
	17791	-1.57	0.11	19512	-0.19	0.22
	19331	0.87	0.17	16558	-0.15	-0.02
	17815	-1.76	0.18	16560	-0.02	-0.08
	14953	-1.43	0.20	17788	0.01	-0.15
	19512	-0.19	0.22	19513	0.12	0.08
	19332	-0.32	0.23	18890	0.18	-0.07
	13890	-2.23	0.24	18891	0.20	-0.18
	14951	-1.68	0.24	19330	0.21	-0.03
	20245	-0.31	0.25	16559	0.50	-0.17
	18842	-2.16	0.28	19331	0.87	0.17

2.7.1.2 Grades 2–3

Table 2.7.1.2

Equating Summary: List 2–3 S602 Online

Table 2.7.1.2

Equating Summary: List 2-3 S602 Online

Comparison of Forms	Form 602		Form 601			
	No. of Items	Average Difficulty (Std. Dev.)	No. of Items	Average Difficulty (Std. Dev.)		
	54	-0.68 (1.57)	54	-0.84 (1.77)		
	Easiest	Hardest	Easiest	Hardest		
	-3.93	2.23	-4.25	2.60		
Anchoring Items	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	29	-0.49 (1.33)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	29	-0.49 (1.33)				
	Percentage Anchors	Average Displacement				
	54%	-0.01				
Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	18894	-1.65	-0.29	12825	-3.26	0.12
	16685	0.53	-0.28	13790	-2.68	0.03
	19350	0.81	-0.28	20299	-2.40	-0.15
	20301	-0.94	-0.20	13910	-2.33	0.26
	19366	-1.11	-0.20	12828	-2.29	-0.18
	12828	-2.29	-0.18	18894	-1.65	-0.29
	20299	-2.40	-0.15	20300	-1.36	-0.08
	12971	0.35	-0.08	12830	-1.17	-0.07
	20300	-1.36	-0.08	16652	-1.17	0.00
	12830	-1.17	-0.07	12957	-1.12	0.18
	19343	0.97	-0.07	19366	-1.11	-0.20
	19344	0.37	-0.05	17771	-1.07	0.21
	19494	1.51	-0.01	20301	-0.94	-0.20
	14883	0.51	0.00	13911	-0.58	0.22
	16652	-1.17	0.00	16686	-0.47	0.00

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	16686	-0.47	0.00	13912	-0.24	0.04
	19352	1.76	0.03	20264	0.28	0.20
	13790	-2.68	0.03	12971	0.35	-0.08
	19351	0.87	0.04	19344	0.37	-0.05
	13912	-0.24	0.04	14883	0.51	0.00
	19492	0.51	0.06	19492	0.51	0.06
	12825	-3.26	0.12	16685	0.53	-0.28
	20266	0.69	0.13	20265	0.55	0.21
	12957	-1.12	0.18	20266	0.69	0.13
	20264	0.28	0.20	19350	0.81	-0.28
	20265	0.55	0.21	19351	0.87	0.04
	17771	-1.07	0.21	19343	0.97	-0.07
	13911	-0.58	0.22	19494	1.51	-0.01
	13910	-2.33	0.26	19352	1.76	0.03

2.7.1.3 Grades 4–5

Table 2.7.1.3

Equating Summary: List 4–5 S602 Online

Table 2.7.1.3

Equating Summary: List 4-5 S602 Online

Comparison of Forms	Form 602		Form 601			
	No. of Items	Average Difficulty (Std. Dev.)	No. of Items	Average Difficulty (Std. Dev.)		
	54	0.72 (1.29)	54	0.66 (1.37)		
	Easiest	Hardest	Easiest	Hardest		
	-2.23	3.80	-2.36	3.33		
Anchoring Items	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	32	0.94 (1.17)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	32	0.94 (1.17)				
	Percentage Anchors	Average Displacement				
	59%	0.00				
Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	14939	2.28	-0.29	19521	-1.16	-0.01
	20269	-0.24	-0.27	18720	-0.52	-0.20
	18720	-0.52	-0.20	16710	-0.52	0.24
	20268	0.06	-0.18	18628	-0.49	-0.04
	16619	2.15	-0.14	19520	-0.24	0.00
	19522	0.38	-0.14	20269	-0.24	-0.27
	17792	0.70	-0.12	18718	-0.11	0.14
	14946	0.95	-0.11	20268	0.06	-0.18
	16714	2.68	-0.10	17789	0.08	-0.06
	18617	2.68	-0.08	14945	0.10	0.19
	17789	0.08	-0.06	16615	0.10	0.14
	18628	-0.49	-0.04	16616	0.29	0.25
	16620	2.50	-0.04	14941	0.32	0.25
	19425	2.18	-0.02	19522	0.38	-0.14
	19370	1.19	-0.02	17790	0.45	-0.01

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	17790	0.45	-0.01	14940	0.48	-0.01
	14940	0.48	-0.01	17792	0.70	-0.12
	19521	-1.16	-0.01	14946	0.95	-0.11
	19520	-0.24	0.00	16709	1.19	0.04
	16709	1.19	0.04	19370	1.19	-0.02
	19426	2.11	0.05	18616	1.21	0.06
	19372	2.32	0.06	16713	1.64	0.10
	18616	1.21	0.06	19426	2.11	0.05
	19424	2.59	0.06	16619	2.15	-0.14
	16713	1.64	0.10	19425	2.18	-0.02
	19371	2.64	0.12	14939	2.28	-0.29
	18718	-0.11	0.14	19372	2.32	0.06
	16615	0.10	0.14	16620	2.50	-0.04
	14945	0.10	0.19	19424	2.59	0.06
	16710	-0.52	0.24	19371	2.64	0.12
	14941	0.32	0.25	18617	2.68	-0.08
	16616	0.29	0.25	16714	2.68	-0.10

2.7.1.4 Grades 6–8

Table 2.7.1.4

Equating Summary: List 6–8 S602 Online

Table 2.7.1.4

Equating Summary: List 6-8 S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Items		Average Difficulty (Std. Dev.)	No. of Items		Average Difficulty (Std. Dev.)
	54		1.20 (1.10)	54		1.19 (1.05)
	Easiest		Hardest	Easiest		Hardest
	-1.14		3.85	-1.14		3.49
Anchoring Items	No. of Possible Anchors		Average Difficulty (Std. Dev.)			
	28		0.91 (0.91)			
	No. of Anchors Used		Average Difficulty (Std. Dev.)			
	28		0.91 (0.91)			
	Percentage Anchors		Average Displacement			
	52%		0.00			
Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	16568	2.04	-0.25	17679	-1.14	0.17
	19445	0.86	-0.24	20074	-0.27	0.07
	20274	0.77	-0.22	18898	-0.23	0.00
	19444	0.67	-0.21	19287	-0.06	-0.03
	14916	1.73	-0.18	20076	-0.02	0.05
	14917	1.05	-0.15	18897	0.00	0.13
	19318	2.04	-0.14	17680	0.05	0.06
	14859	1.51	-0.07	16664	0.10	0.13
	20075	0.74	-0.04	20272	0.30	0.03
	16566	1.86	-0.04	19444	0.67	-0.21
	19287	-0.06	-0.03	19286	0.71	0.02
	19319	1.56	-0.02	20075	0.74	-0.04
	18898	-0.23	0.00	20274	0.77	-0.22
	16665	1.13	0.01	19445	0.86	-0.24
	14858	1.54	0.01	14917	1.05	-0.15

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	19286	0.71	0.02	14915	1.08	0.22
	16666	1.39	0.02	16665	1.13	0.01
	20272	0.30	0.03	19320	1.34	0.13
	20076	-0.02	0.05	16666	1.39	0.02
	17680	0.05	0.06	14859	1.51	-0.07
	20074	-0.27	0.07	14858	1.54	0.01
	19320	1.34	0.13	19319	1.56	-0.02
	18897	0.00	0.13	20078	1.62	0.25
	16664	0.10	0.13	14916	1.73	-0.18
	17679	-1.14	0.17	16566	1.86	-0.04
	14915	1.08	0.22	16568	2.04	-0.25
	20078	1.62	0.25	19318	2.04	-0.14
	16567	3.07	0.28	16567	3.07	0.28

2.7.1.5 Grades 9–12

Table 2.7.1.5

Equating Summary: List 9–12 Online

Table 2.7.1.5

Equating Summary: List 9-12 S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Items		Average Difficulty (Std. Dev.)	No. of Items		Average Difficulty (Std. Dev.)
	54		1.87 (1.14)	54		1.66 (1.10)
	Easiest		Hardest	Easiest		Hardest
	-0.48		4.08	-0.48		4.08
Anchoring Items	No. of Possible Anchors		Average Difficulty (Std. Dev.)			
	36		1.65 (1.14)			
	No. of Anchors Used		Average Difficulty (Std. Dev.)			
	36		1.65 (1.14)			
	Percentage Anchors		Average Displacement			
	67%		0.01			
Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	17755	4.08	-0.26	18573	-0.48	-0.09
	16658	2.48	-0.25	17761	-0.38	0.00
	16656	2.18	-0.23	19310	-0.14	0.09
	17749	2.88	-0.20	17719	0.10	-0.10
	17754	2.67	-0.18	18574	0.26	0.13
	17721	2.25	-0.14	18566	0.28	-0.08
	20094	0.46	-0.13	20094	0.46	-0.13
	17750	1.98	-0.13	18565	0.53	0.18
	20325	2.69	-0.12	20323	0.59	0.23
	17719	0.10	-0.10	19302	0.77	-0.02
	16657	1.04	-0.10	19311	0.79	0.15
	18573	-0.48	-0.09	17762	0.94	-0.06
	17753	2.10	-0.09	16657	1.04	-0.10
	18566	0.28	-0.08	17720	1.18	0.19
	17763	1.62	-0.08	18567	1.48	0.16

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	19290	2.86	-0.08	17763	1.62	-0.08
	17762	0.94	-0.06	20095	1.77	0.05
	20324	2.05	-0.02	17750	1.98	-0.13
	19302	0.77	-0.02	20324	2.05	-0.02
	20096	2.38	-0.02	20319	2.05	0.01
	17761	-0.38	0.00	17753	2.10	-0.09
	20319	2.05	0.01	20233	2.18	0.22
	20095	1.77	0.05	16656	2.18	-0.23
	19292	2.93	0.09	17721	2.25	-0.14
	19310	-0.14	0.09	20231	2.25	0.28
	18574	0.26	0.13	20096	2.38	-0.02
	19311	0.79	0.15	16658	2.48	-0.25
	18567	1.48	0.16	17754	2.67	-0.18
	18565	0.53	0.18	20325	2.69	-0.12
	19358	2.70	0.18	19358	2.70	0.18
	17720	1.18	0.19	19290	2.86	-0.08
	20233	2.18	0.22	17749	2.88	-0.20
	20232	3.03	0.23	19292	2.93	0.09
	20323	0.59	0.23	19360	3.00	0.43
	20231	2.25	0.28	20232	3.03	0.23
	19360	3.00	0.43	17755	4.08	-0.26

2.7.2 Reading

2.7.2.1 Grade 1

Table 2.7.2.1

Equating Summary: Read 1 S602 Online

Table 2.7.2.1

Equating Summary: Read 1 S602 Online

Comparison of Forms	Form 602		Form 601			
	No. of Items	Average Difficulty (Std. Dev.)	No. of Items	Average Difficulty (Std. Dev.)		
	66	-0.96 (1.12)	72	-0.98 (0.99)		
	Easiest	Hardest	Easiest	Hardest		
	-4.24	0.84	-3.60	0.84		
Anchoring Items	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	38	-0.87 (0.93)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	38	-0.87 (0.93)				
	Percentage Anchors	Average Displacement				
	58%	-0.02				
Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	18450	-0.95	-0.27	20340	-3.44	-0.20
	20339	0.06	-0.26	20102	-2.56	0.12
	18467	-2.16	-0.23	17954	-2.20	-0.06
	17959	0.18	-0.23	18467	-2.16	-0.23
	13195	-1.52	-0.20	18465	-2.12	-0.20
	18465	-2.12	-0.20	13193	-2.11	0.12
	20340	-3.44	-0.20	13194	-2.06	0.08
	17983	-1.16	-0.17	20407	-2.03	-0.14
	17982	-0.38	-0.17	13195	-1.52	-0.20
	20407	-2.03	-0.14	18466	-1.28	-0.06
	18098	0.02	-0.12	20104	-1.19	0.08
	17131	-0.34	-0.12	17983	-1.16	-0.17
	18099	0.46	-0.09	17984	-1.02	0.01
	18100	0.43	-0.09	18539	-0.96	0.22
	17954	-2.20	-0.06	19387	-0.95	0.00

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	18466	-1.28	-0.06	18450	-0.95	-0.27
	17956	-0.74	-0.05	19626	-0.76	0.21
	17958	-0.34	-0.04	17956	-0.74	-0.05
	17133	0.84	-0.04	17986	-0.74	0.03
	17960	-0.19	-0.01	19624	-0.62	0.15
	19387	-0.95	0.00	20103	-0.61	0.02
	17984	-1.02	0.01	19634	-0.55	0.26
	20103	-0.61	0.02	19389	-0.54	0.16
	18538	-0.25	0.02	17955	-0.54	0.08
	17986	-0.74	0.03	17132	-0.43	0.25
	17955	-0.54	0.08	17982	-0.38	-0.17
	20104	-1.19	0.08	17131	-0.34	-0.12
	13194	-2.06	0.08	17958	-0.34	-0.04
	19632	-0.30	0.09	19632	-0.30	0.09
	13193	-2.11	0.12	18538	-0.25	0.02
	20102	-2.56	0.12	17960	-0.19	-0.01
	19624	-0.62	0.15	18098	0.02	-0.12
	19389	-0.54	0.16	20339	0.06	-0.26
	17987	0.09	0.20	17987	0.09	0.20
	19626	-0.76	0.21	17959	0.18	-0.23
	18539	-0.96	0.22	18100	0.43	-0.09
	17132	-0.43	0.25	18099	0.46	-0.09
	19634	-0.55	0.26	17133	0.84	-0.04

2.7.2.2 Grades 2–3

Table 2.7.2.2

Equating Summary: Read 2–3 S602 Online

Table 2.7.2.2

Equating Summary: Read 2-3 S602 Online

Comparison of Forms	Form 602		Form 601			
	No. of Items	Average Difficulty (Std. Dev.)	No. of Items	Average Difficulty (Std. Dev.)		
	69	0.08 (0.97)	72	0.08 (0.83)		
	Easiest	Hardest	Easiest	Hardest		
	-2.07	2.46	-1.95	2.46		
Anchoring Items	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	44	0.23 (0.79)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	44	0.23 (0.79)				
	Percentage Anchors	Average Displacement				
	64%	-0.02				
Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	18363	0.11	-0.29	17887	-1.41	-0.07
	19391	-0.41	-0.27	17879	-1.41	-0.17
	19403	0.72	-0.26	17888	-0.96	-0.14
	18475	0.16	-0.25	17886	-0.83	0.12
	18361	-0.67	-0.23	19404	-0.82	0.18
	19405	0.50	-0.20	19401	-0.79	0.12
	13345	1.24	-0.17	18361	-0.67	-0.23
	16092	1.27	-0.17	17880	-0.52	-0.14
	17892	0.42	-0.17	19391	-0.41	-0.27
	17879	-1.41	-0.17	20413	-0.27	0.18
	18473	0.01	-0.16	19575	-0.26	0.23
	17880	-0.52	-0.14	13340	-0.25	0.03
	17888	-0.96	-0.14	20368	-0.18	0.11
	13346	0.60	-0.13	19574	-0.16	0.26
	16095	0.70	-0.12	20369	-0.13	-0.01

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	17894	0.23	-0.11	18473	0.01	-0.16
	13344	0.33	-0.10	20414	0.02	0.08
	13339	0.38	-0.09	17893	0.05	0.17
	17049	1.22	-0.08	18363	0.11	-0.29
	18474	0.25	-0.08	18475	0.16	-0.25
	17050	1.25	-0.08	17894	0.23	-0.11
	17887	-1.41	-0.07	18474	0.25	-0.08
	17928	2.46	-0.07	17051	0.32	0.18
	20415	0.38	-0.07	13344	0.33	-0.10
	13338	0.80	-0.05	20415	0.38	-0.07
	20369	-0.13	-0.01	13339	0.38	-0.09
	13340	-0.25	0.03	19573	0.39	0.24
	16094	0.90	0.03	17892	0.42	-0.17
	20367	0.61	0.06	19405	0.50	-0.20
	20414	0.02	0.08	13346	0.60	-0.13
	20368	-0.18	0.11	20367	0.61	0.06
	19401	-0.79	0.12	18366	0.64	0.21
	17886	-0.83	0.12	16095	0.70	-0.12
	17893	0.05	0.17	19403	0.72	-0.26
	17051	0.32	0.18	13338	0.80	-0.05
	19404	-0.82	0.18	16094	0.90	0.03
	20413	-0.27	0.18	19652	0.98	0.22
	18366	0.64	0.21	17924	1.13	0.26
	19652	0.98	0.22	17049	1.22	-0.08
	19575	-0.26	0.23	13345	1.24	-0.17
	19573	0.39	0.24	17050	1.25	-0.08
	17924	1.13	0.26	16092	1.27	-0.17
	19574	-0.16	0.26	17934	1.31	0.26
	17934	1.31	0.26	17928	2.46	-0.07

2.7.2.3 Grades 4–5

Table 2.7.2.3

Equating Summary: Read 4–5 S602 Online

Table 2.7.2.3

Equating Summary: Read 4-5 S602 Online

Comparison of Forms	Form 602		Form 601			
	No. of Items	Average Difficulty (Std. Dev.)	No. of Items	Average Difficulty (Std. Dev.)		
	72	1.03 (1.06)	72	0.94 (1.08)		
	Easiest	Hardest	Easiest	Hardest		
	-2.04	2.99	-2.04	2.99		
Anchoring Items	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	55	1.06 (1.06)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	55	1.06 (1.06)				
	Percentage Anchors	Average Displacement				
	76%	0.00				
Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	16019	1.59	-0.26	20118	-2.04	-0.03
	16010	1.34	-0.26	13407	-0.72	-0.06
	19761	1.59	-0.25	18409	-0.65	0.02
	18198	0.71	-0.23	18184	-0.52	0.14
	15708	1.38	-0.21	20115	-0.38	-0.04
	18485	1.89	-0.21	20119	-0.26	-0.07
	17110	1.17	-0.19	13409	-0.17	0.10
	18413	1.14	-0.17	16009	-0.10	0.10
	19762	0.70	-0.17	18410	-0.06	-0.07
	18185	1.15	-0.16	17109	-0.03	0.19
	18186	0.66	-0.14	13408	0.06	0.20
	18487	2.15	-0.14	19525	0.07	-0.10
	16017	0.09	-0.14	16017	0.09	-0.14
	16011	0.12	-0.13	20116	0.11	-0.10
	20120	0.49	-0.13	16011	0.12	-0.13

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	15706	0.21	-0.11	15706	0.21	-0.11
	20116	0.11	-0.10	20120	0.49	-0.13
	19525	0.07	-0.10	20128	0.61	0.05
	18486	2.40	-0.09	18186	0.66	-0.14
	18416	1.73	-0.08	19762	0.70	-0.17
	18410	-0.06	-0.07	18198	0.71	-0.23
	20119	-0.26	-0.07	18197	0.80	-0.02
	19589	1.27	-0.06	18125	0.98	0.12
	13407	-0.72	-0.06	18128	0.99	0.24
	20115	-0.38	-0.04	18413	1.14	-0.17
	18196	1.36	-0.04	18185	1.15	-0.16
	20118	-2.04	-0.03	17110	1.17	-0.19
	20446	1.76	-0.02	15707	1.23	0.10
	18197	0.80	-0.02	19589	1.27	-0.06
	18123	1.39	0.00	16018	1.28	0.03
	18415	2.44	0.01	16010	1.34	-0.26
	18409	-0.65	0.02	18196	1.36	-0.04
	16018	1.28	0.03	15708	1.38	-0.21
	19590	2.27	0.03	18123	1.39	0.00
	20128	0.61	0.05	17111	1.47	0.06
	17111	1.47	0.06	19761	1.59	-0.25
	20127	2.41	0.08	16019	1.59	-0.26
	16009	-0.10	0.10	19757	1.68	0.20
	15707	1.23	0.10	20126	1.71	0.26
	13409	-0.17	0.10	18416	1.73	-0.08
	18125	0.98	0.12	20446	1.76	-0.02
	19758	2.54	0.13	20170	1.77	0.19
	18184	-0.52	0.14	20445	1.80	0.20
	20447	2.73	0.15	18485	1.89	-0.21
	19759	2.68	0.16	20284	1.94	0.23

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	17109	-0.03	0.19	18487	2.15	-0.14
	20170	1.77	0.19	19590	2.27	0.03
	19757	1.68	0.20	20172	2.33	0.29
	13408	0.06	0.20	18486	2.40	-0.09
	20445	1.80	0.20	20127	2.41	0.08
	20171	2.99	0.22	18415	2.44	0.01
	20284	1.94	0.23	19758	2.54	0.13
	18128	0.99	0.24	19759	2.68	0.16
	20126	1.71	0.26	20447	2.73	0.15
	20172	2.33	0.29	20171	2.99	0.22

2.7.2.4 Grades 6–8

Table 2.7.2.4

Equating Summary: Read 6–8 S602 Online

Table 2.7.2.4

Equating Summary: Read 6-8 S602 Online

Comparison of Forms	Form 602		Form 601			
	No. of Items	Average Difficulty (Std. Dev.)	No. of Items	Average Difficulty (Std. Dev.)		
	69	1.50 (1.32)	72	1.48 (1.31)		
	Easiest	Hardest	Easiest	Hardest		
	-1.36	3.82	-1.69	3.79		
Anchoring Items	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	34	1.66 (1.29)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	34	1.66 (1.29)				
	Percentage Anchors	Average Displacement				
	49%	0.04				
Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	18507	2.47	-0.28	19616	-0.77	0.21
	19474	0.75	-0.19	19472	-0.67	0.10
	20215	1.11	-0.18	19617	-0.60	0.03
	19473	0.38	-0.18	18321	-0.28	0.21
	18506	1.58	-0.18	18381	0.22	0.04
	19484	2.37	-0.17	19473	0.38	-0.18
	13629	0.78	-0.12	19474	0.75	-0.19
	19700	2.13	-0.09	13629	0.78	-0.12
	18055	1.14	-0.08	13631	0.84	0.16
	18056	1.17	-0.04	19485	0.86	0.00
	20138	0.88	-0.01	20138	0.88	-0.01
	19485	0.86	0.00	20140	0.99	0.05
	20388	2.25	0.01	20215	1.11	-0.18
	13630	1.47	0.03	18055	1.14	-0.08
	19617	-0.60	0.03	18056	1.17	-0.04

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	18054	1.38	0.03	18054	1.38	0.03
	20454	2.83	0.03	13630	1.47	0.03
	18381	0.22	0.04	18506	1.58	-0.18
	20139	1.62	0.04	20139	1.62	0.04
	19486	2.98	0.04	19700	2.13	-0.09
	20140	0.99	0.05	20388	2.25	0.01
	19472	-0.67	0.10	19484	2.37	-0.17
	20208	3.27	0.13	18507	2.47	-0.28
	20455	2.91	0.14	19701	2.58	0.15
	19701	2.58	0.15	20209	2.69	0.24
	13631	0.84	0.16	20454	2.83	0.03
	19616	-0.77	0.21	20455	2.91	0.14
	18321	-0.28	0.21	19486	2.98	0.04
	20207	3.79	0.22	20457	3.06	0.28
	20458	3.41	0.24	19702	3.15	0.25
	20209	2.69	0.24	20208	3.27	0.13
	19702	3.15	0.25	20458	3.41	0.24
	20459	3.59	0.27	20459	3.59	0.27
	20457	3.06	0.28	20207	3.79	0.22

2.7.2.5 Grades 9–12

Table 2.7.2.5

Equating Summary: Read 9–12 S602 Online

Table 2.7.2.5

Equating Summary: Read 9-12 S602 Online

Comparison of Forms	Form 602		Form 601			
	No. of Items	Average Difficulty (Std. Dev.)	No. of Items	Average Difficulty (Std. Dev.)		
	72	2.37 (1.23)	72	2.34 (1.31)		
	Easiest	Hardest	Easiest	Hardest		
	-0.27	4.67	-1.20	4.52		
Anchoring Items	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	45	2.49 (1.19)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	45	2.49 (1.19)				
	Percentage Anchors	Average Displacement				
	63%	0.01				
Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	20347	2.65	-0.30	20199	-0.26	-0.13
	20146	0.19	-0.29	20146	0.19	-0.29
	19464	2.96	-0.26	18509	0.26	0.11
	17077	4.52	-0.23	17998	0.45	0.06
	16072	3.26	-0.18	17996	0.59	0.16
	18519	3.06	-0.18	20182	1.04	0.01
	19672	3.11	-0.15	19660	1.09	-0.10
	20200	1.50	-0.15	18446	1.10	0.12
	20199	-0.26	-0.13	20148	1.41	-0.07
	18518	2.79	-0.13	20147	1.48	-0.05
	19466	3.96	-0.12	20200	1.50	-0.15
	20154	3.01	-0.10	18455	1.53	-0.04
	19660	1.09	-0.10	18517	1.58	0.12
	20461	3.19	-0.09	18025	1.88	0.18
	16070	2.65	-0.08	18024	2.25	0.22

Displacement of Anchor Items	Anchor Items by Displacement			Anchor Items by Item Difficulty		
	Item ID	Item Difficulty	Displacement	Item ID	Item Difficulty	Displacement
	20148	1.41	-0.07	18030	2.25	0.13
	20147	1.48	-0.05	16071	2.36	0.09
	19596	3.25	-0.05	17076	2.40	0.19
	18455	1.53	-0.04	19465	2.41	0.19
	19673	3.14	-0.02	20347	2.65	-0.30
	20182	1.04	0.01	16070	2.65	-0.08
	18527	2.95	0.01	20183	2.66	0.07
	18526	3.40	0.01	18518	2.79	-0.13
	20184	2.82	0.03	20184	2.82	0.03
	19597	3.56	0.03	18527	2.95	0.01
	18032	3.45	0.04	19464	2.96	-0.26
	17998	0.45	0.06	20154	3.01	-0.10
	20183	2.66	0.07	18519	3.06	-0.18
	18525	4.09	0.08	19672	3.11	-0.15
	16071	2.36	0.09	19673	3.14	-0.02
	19598	3.82	0.11	20461	3.19	-0.09
	18509	0.26	0.11	19596	3.25	-0.05
	18517	1.58	0.12	16072	3.26	-0.18
	18446	1.10	0.12	20352	3.35	0.24
	18031	3.57	0.13	18526	3.40	0.01
	18030	2.25	0.13	18032	3.45	0.04
	17996	0.59	0.16	19597	3.56	0.03
	19674	3.97	0.17	18031	3.57	0.13
	18025	1.88	0.18	20462	3.60	0.28
	19465	2.41	0.19	19598	3.82	0.11
	17076	2.40	0.19	19466	3.96	-0.12
	18024	2.25	0.22	19674	3.97	0.17
	20352	3.35	0.24	20351	3.99	0.27
	20351	3.99	0.27	18525	4.09	0.08
	20462	3.60	0.28	17077	4.52	-0.23

2.7.3 Writing

2.7.3.1 Grade 1

Table 2.7.3.1.1

Equating Summary: Writ 1 A S602 Online

Table 2.7.3.1.1

Equating Summary: Writ 1 A S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks	Average Difficulty (Std. Dev.)	No. of Tasks	Average Difficulty (Std. Dev.)		
	2	-0.50 (0.37)	2	-0.41 (0.24)		
	Easiest	Hardest	Easiest	Hardest		
	-0.76	-0.24	-0.58	-0.24		
Anchoring Tasks	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	2	-0.50 (0.37)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	2	-0.50 (0.37)				
	Percentage Anchors	Average Displacement				
	100%	0.00				
Rating Scale Step Measures by Task	Anchored Scale Steps					
	Step	Measure				
	1	-2.47				
	2	-2.78				
	3	-2.61				
	4	-1.68				
	5	-0.48				
	6	0.97				
	7	2.25				
	8	3.21				
	9	3.59				
Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty		
	Task ID	Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
	20506	-0.76	0.00	20506	-0.76	0.00
	19805	-0.24	0.00	19805	-0.24	0.00

Table 2.7.3.1.2

Equating Summary: Writ 1 B/C S602 Online

Table 2.7.3.1.2

Equating Summary: Writ 1 B/C S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks		Average Difficulty (Std. Dev.)	No. of Tasks		Average Difficulty (Std. Dev.)
	2		0.16 (0.35)	2		-0.20 (0.16)
	Easiest		Hardest	Easiest		Hardest
	-0.09		0.41	-0.31		-0.09
Anchoring Tasks	No. of Possible Anchors		Average Difficulty (Std. Dev.)			
	2		0.16 (0.35)			
	No. of Anchors Used		Average Difficulty (Std. Dev.)			
	2		0.16 (0.35)			
	Percentage Anchors		Average Displacement			
	100%		0.00			
Common Rating Scale Step Measures	Anchored Scale Steps					
	Step		Measure			
	1		-2.47			
	2		-2.78			
	3		-2.61			
	4		-1.68			
	5		-0.48			
	6		0.97			
	7		2.25			
	8		3.21			
	9		3.59			
Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty		
	Task ID	Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
	19827	-0.09	-0.03	19827	-0.09	-0.03
	20511	0.41	0.02	20511	0.41	0.02

2.7.3.2 Grades 2–3

Table 2.7.3.2.1

Equating Summary: Writ 2–3 A S602 Online

Table 2.7.3.2.1

Equating Summary: Writ 2-3 A S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks		Average Difficulty (Std. Dev.)	No. of Tasks		Average Difficulty (Std. Dev.)
	2		0.03 (0.06)	2		-0.05 (0.05)
	Easiest		Hardest	Easiest		Hardest
	-0.02		0.07	-0.09		-0.02
Anchoring Tasks	No. of Possible Anchors		Average Difficulty (Std. Dev.)			
	2		0.03 (0.06)			
	No. of Anchors Used		Average Difficulty (Std. Dev.)			
	2		0.03 (0.06)			
	Percentage Anchors		Average Displacement			
	100%		0.00			
Common Rating Scale Step Measures	Anchored Scale Steps					
	Step		Measure			
	1		-2.47			
	2		-2.78			
	3		-2.61			
	4		-1.68			
	5		-0.48			
	6		0.97			
	7		2.25			
	8		3.21			
	9		3.59			
	Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty	
Task ID		Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
19808		-0.02	-0.03	19808	-0.02	-0.03
20543		0.07	0.03	20543	0.07	0.03

Table 2.7.3.2.2**Equating Summary: Writ 2–3 B/C S602 Online**

Table 2.7.3.2.2

Equating Summary: Writ 2-3 B/C S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks		Average Difficulty (Std. Dev.)	No. of Tasks		Average Difficulty (Std. Dev.)
	2		0.28 (0.05)	2		0.28 (0.06)
	Easiest		Hardest	Easiest		Hardest
	0.25		0.31	0.25		0.32
Anchoring Tasks	No. of Possible Anchors		Average Difficulty (Std. Dev.)			
	2		0.28 (0.05)			
	No. of Anchors Used		Average Difficulty (Std. Dev.)			
	2		0.28 (0.05)			
	Percentage Anchors		Average Displacement			
	100%		0.00			
Common Rating Scale Step Measures	Anchored Scale Steps					
	Step		Measure			
	1		-2.47			
	2		-2.78			
	3		-2.61			
	4		-1.68			
	5		-0.48			
	6		0.97			
	7		2.25			
	8		3.21			
	9		3.59			
	Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty	
Task ID		Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
19829		0.25	-0.05	19829	0.25	-0.05
20541		0.31	0.05	20541	0.31	0.05

2.7.3.3 Grades 4–5

Table 2.7.3.3.1

Equating Summary: Writ 4–5 A S602 Online

Table 2.7.3.3.1

Equating Summary: Writ 4-5 A S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks		Average Difficulty (Std. Dev.)	No. of Tasks		Average Difficulty (Std. Dev.)
	2		1.15 (0.13)	2		1.13 (0.17)
	Easiest		Hardest	Easiest		Hardest
	1.05		1.24	1.01		1.24
Anchoring Tasks	No. of Possible Anchors		Average Difficulty (Std. Dev.)			
	2		1.15 (0.13)			
	No. of Anchors Used		Average Difficulty (Std. Dev.)			
	2		1.15 (0.13)			
	Percentage Anchors		Average Displacement			
	100%		0.00			
Common Rating Scale Step Measures	Anchored Scale Steps					
	Step		Measure			
	1		-2.47			
	2		-2.78			
	3		-2.61			
	4		-1.68			
	5		-0.48			
	6		0.97			
	7		2.25			
	8		3.21			
	9		3.59			
Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty		
	Task ID	Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
	20509	1.05	-0.02	20509	1.05	-0.02
	19814_20289	1.24	0.02	19814_20289	1.24	0.02

Table 2.7.3.3.2

Equating Summary: Writ 4–5 B/C S602 Online

Table 2.7.3.3.2

Equating Summary: Writ 4-5 B/C S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks	Average Difficulty (Std. Dev.)	No. of Tasks	Average Difficulty (Std. Dev.)		
	2	2.04 (0.08)	2	2.25 (0.38)		
	Easiest	Hardest	Easiest	Hardest		
1.98	2.10	1.98	2.52			
Anchoring Tasks	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	2	2.04 (0.08)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	2	2.04 (0.08)				
	Percentage Anchors	Average Displacement				
	100%	0.00				
Common Rating Scale Step Measures	Anchored Scale Steps					
	Step	Measure				
	1	-2.47				
	2	-2.78				
	3	-2.61				
	4	-1.68				
	5	-0.48				
	6	0.97				
	7	2.25				
	8	3.21				
	9	3.59				
Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty		
	Task ID	Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
	20518	2.10	-0.06	19833_20219	1.98	0.06
	19833_20219	1.98	0.06	20518	2.10	-0.06

2.7.3.4 Grades 6–8

Table 2.7.3.4.1

Equating Summary: Writ 6–8 A S602 Online

Table 2.7.3.4.1

Equating Summary: Writ 6–8 A S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks	Average Difficulty (Std. Dev.)	No. of Tasks	Average Difficulty (Std. Dev.)		
	2	0.86 (0.07)	2	0.76 (0.22)		
	Easiest	Hardest	Easiest	Hardest		
	0.82	0.91	0.60	0.91		
Anchoring Tasks	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	2	0.86 (0.07)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	2	0.86 (0.07)				
	Percentage Anchors	Average Displacement				
	100%	0.00				
Common Rating Scale Step Measures	Anchored Scale Steps					
	Step	Measure				
	1	-2.47				
	2	-2.78				
	3	-2.61				
	4	-1.68				
	5	-0.48				
	6	0.97				
	7	2.25				
	8	3.21				
	9	3.59				
	Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty	
Task ID		Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
19817_20399		0.91	-0.01	20605	0.82	0.01
20605		0.82	0.01	19817_20399	0.91	-0.01

Table 2.7.3.4.2

Equating Summary: Writ 6–8 B/C S602 Online

Table 2.7.3.4.2

Equating Summary: Writ 6-8 B/C S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks	Average Difficulty (Std. Dev.)	No. of Tasks	Average Difficulty (Std. Dev.)		
	2	1.50 (0.07)	2	1.26 (0.27)		
	Easiest	Hardest	Easiest	Hardest		
	1.45	1.55	1.07	1.45		
Anchoring Tasks	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	2	1.50 (0.07)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	2	1.50 (0.07)				
	Percentage Anchors	Average Displacement				
	100%	0.00				
Common Rating Scale Step Measures	Anchored Scale Steps					
	Step	Measure				
	1	-2.47				
	2	-2.78				
	3	-2.61				
	4	-1.68				
	5	-0.48				
	6	0.97				
	7	2.25				
	8	3.21				
	9	3.59				
Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty		
	Task ID	Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
	20610	1.55	-0.01	19837_20225	1.45	0.01
	19837_20225	1.45	0.01	20610	1.55	-0.01

2.7.3.5 Grades 9–12

Table 2.7.3.5.1

Equating Summary: Writ 9–12 A S602 Online

Table 2.7.3.5.1

Equating Summary: Writ 9-12 A S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks		Average Difficulty (Std. Dev.)	No. of Tasks		Average Difficulty (Std. Dev.)
	2		2.01 (0.22)	2		2.12 (0.07)
	Easiest		Hardest	Easiest		Hardest
	1.86		2.17	2.06		2.17
Anchoring Tasks	No. of Possible Anchors		Average Difficulty (Std. Dev.)			
	2		2.01 (0.22)			
	No. of Anchors Used		Average Difficulty (Std. Dev.)			
	2		2.01 (0.22)			
	Percentage Anchors		Average Displacement			
	100%		0.00			
Common Rating Scale Step Measures	Anchored Scale Steps					
	Step		Measure			
	1		-2.47			
	2		-2.78			
	3		-2.61			
	4		-1.68			
	5		-0.48			
	6		0.97			
	7		2.25			
	8		3.21			
	9		3.59			
	Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty	
Task ID		Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
19786_20396		2.17	-0.01	20501	1.86	0.01
20501		1.86	0.01	19786_20396	2.17	-0.01

Table 2.7.3.5.2

Equating Summary: Writ 9–12 B/C S602 Online

Table 2.7.3.5.2

Equating Summary: Writ 9-12 B/C S602 Online

Comparison of Forms*	Form 602			Form 601		
	No. of Tasks	Average Difficulty (Std. Dev.)	No. of Tasks	Average Difficulty (Std. Dev.)		
	2	2.00 (0.45)	2	2.00 (0.45)		
	Easiest	Hardest	Easiest	Hardest		
1.68	2.32	1.00	2.32			
Anchoring Tasks	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	1	2.32 (N/A)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	1	2.32 (N/A)				
	Percentage Anchors	Average Displacement				
	50%	0.00				
Common Rating Scale Step Measures	Anchored Scale Steps					
	Step	Measure				
	1	-2.47				
	2	-2.78				
	3	-2.61				
	4	-1.68				
	5	-0.48				
	6	0.97				
	7	2.25				
	8	3.21				
	9	3.59				
Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty		
	Task ID	Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
	17319 18252	2.32	0.00	17319 18252	2.32	0.00

2.7.4 Speaking

2.7.4.1 Grade 1

Table 2.7.4.1

Equating Summary: Spek 1 S602 Online

Table 2.7.4.1

Equating Summary: Spek 1 S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks	Average Difficulty (Std. Dev.)		No. of Tasks	Average Difficulty (Std. Dev.)	
	9	-1.71 (2.02)		9	-1.71 (2.27)	
	Easiest	Hardest		Easiest	Hardest	
	-4.76	0.11		-4.76	0.61	
Anchoring Tasks	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	3	-2.06 (2.35)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	3	-2.06 (2.35)				
	Percentage Anchors	Average Displacement				
	33%	0.13				
Rating Scale Step Measures by Task	Anchored Scale Steps					
	Task	Step	Measure			
	PL 1 Tasks	1	0.56			
		2	-0.56			
	PL 3/PL 5 Tasks	1	-2.65			
		2	-1.80			
		3	1.46			
		4	2.98			
Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty		
	Task ID	Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
	19876	-0.54	-0.08	19864	-4.76	0.51
	19870	-0.87	-0.04	19870	-0.87	-0.04
	19864	-4.76	0.51	19876	-0.54	-0.08

2.7.4.2 Grades 2–3

Table 2.7.4.2

Equating Summary: Spek 2–3 S602 Online

Table 2.7.4.2

Equating Summary: Spek 2-3 S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks		Average Difficulty (Std. Dev.)	No. of Tasks		Average Difficulty (Std. Dev.)
	9		-1.27 (2.39)	9		-1.43 (2.53)
	Easiest		Hardest	Easiest		Hardest
	-4.62		1.22	-4.95		0.66
Anchoring Tasks	No. of Possible Anchors		Average Difficulty (Std. Dev.)			
	3		-1.31 (2.87)			
	No. of Anchors Used		Average Difficulty (Std. Dev.)			
	3		-1.31 (2.87)			
	Percentage Anchors		Average Displacement			
	33%		0.19			
Rating Scale Step Measures by Task	Anchored Scale Steps					
	Task	Step	Measure			
	PL 1 Tasks	1	0.56			
		2	-0.56			
	PL 3/PL 5 Tasks	1	-2.65			
		2	-1.80			
		3	1.46			
		4	2.98			
Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty		
	Task ID	Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
	19885	0.45	-0.08	19144	-4.62	0.60
	19892	0.24	0.04	19892	0.24	0.04
	19144	-4.62	0.60	19885	0.45	-0.08

2.7.4.3 Grades 4–5

Table 2.7.4.3

Equating Summary: Spek 4–5 S602 Online

Table 2.7.4.3

Equating Summary: Spek 4-5 S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks	Average Difficulty (Std. Dev.)		No. of Tasks	Average Difficulty (Std. Dev.)	
	9	0.08 (2.75)		9	-0.22 (2.83)	
	Easiest	Hardest		Easiest	Hardest	
	-3.96	2.61		-4.02	2.35	
Anchoring Tasks	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	3	-0.13 (3.37)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	3	-0.13 (3.37)				
	Percentage Anchors	Average Displacement				
	33%	0.20				
Rating Scale Step Measures by Task	Anchored Scale Steps					
	Task	Step	Measure			
	PL 1 Tasks	1	0.56			
		2	-0.56			
	PL 3/PL 5 Tasks	1	-2.65			
		2	-1.80			
		3	1.46			
		4	2.98			
Displacement of Anchor Tasks	Anchor Tasks by Displacement			Anchor Tasks by Task Difficulty		
	Task ID	Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
	19961	2.35	-0.03	19910	-3.96	0.62
	19954	1.23	0.00	19954	1.23	0.00
	19910	-3.96	0.62	19961	2.35	-0.03

2.7.4.4 Grades 6–8

Table 2.7.4.4

Equating Summary: Spek 6–8 S602 Online

Table 2.7.4.4

Equating Summary: Spek 6-8 S602 Online

Comparison of Forms	Form 602			Form 601				
	No. of Tasks		Average Difficulty (Std. Dev.)	No. of Tasks		Average Difficulty (Std. Dev.)		
	9		0.11 (2.39)	9		0.20 (2.61)		
	Easiest		Hardest	Easiest		Hardest		
	-3.13		2.18	-3.42		2.42		
Anchoring Tasks	No. of Possible Anchors		Average Difficulty (Std. Dev.)					
	3							
	No. of Anchors Used						Average Difficulty (Std. Dev.)	
	3							
	Percentage Anchors							Average Displacement
	33%							
Rating Scale Step Measures by Task	Anchored Scale Steps							
	Task	Step	Measure					
	PL 1 Tasks	1	0.56					
		2	-0.56					
	PL 3/PL 5 Tasks	1	-2.65					
		2	-1.80					
		3	1.46					
		4	2.98					
	Displacement of Anchor Tasks	Anchor Tasks by Displacement					Anchor Tasks by Task Difficulty	
Task ID		Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement		
19985		1.47	-0.18	19980	-3.13	0.28		
19992		2.02	-0.09	19985	1.47	-0.18		
19980		-3.13	0.28	19992	2.02	-0.09		

2.7.4.5 Grades 9–12

Table 2.7.4.5

Equating Summary: Spek 9–12 S602 Online

Table 2.7.4.5

Equating Summary: Spek 9-12 S602 Online

Comparison of Forms	Form 602			Form 601		
	No. of Tasks	Average Difficulty (Std. Dev.)		No. of Tasks	Average Difficulty (Std. Dev.)	
	9	0.22 (2.37)		9	0.37 (2.51)	
	Easiest	Hardest		Easiest	Hardest	
	-3.06	2.24		-3.08	2.88	
Anchoring Tasks	No. of Possible Anchors	Average Difficulty (Std. Dev.)				
	3	0.34 (2.94)				
	No. of Anchors Used	Average Difficulty (Std. Dev.)				
	3	0.34 (2.94)				
	Percentage Anchors	Average Displacement				
	33%	0.04				
Rating Scale Step Measures by Task	Anchored Scale Steps					
	Task	Step	Measure			
	PL 1 Tasks	1	0.56			
		2	-0.56			
	PL 3/PL 5 Tasks	1	-2.65			
		2	-1.80			
		3	1.46			
		4	2.98			
	Displacement of Anchor Tasks	Anchor Tasks by Displacement				
Task ID		Task Difficulty	Displacement	Task ID	Task Difficulty	Displacement
20029		1.98	-0.11	20023	-3.06	0.02
20023		-3.06	0.02	20029	1.98	-0.11
20036		2.09	0.21	20036	2.09	0.21

2.8 Test Characteristic Curve

Test characteristic curves (TCC) graphically show the functional relationship between a student's ability measure (in logits) on the horizontal axis and that student's expected raw score (i.e., the estimated true score) on the vertical axis. Thus, for a given ability measure, the corresponding expected raw score can be found via the TCC. For reporting purposes, WIDA uses the student's ability measure to determine the proficiency level. Since the TCC transforms ability measures to expected raw scores, this representation allows test users to relate a student's ability measure to his/her proficiency level (i.e., a more familiar frame of reference that test users employ to interpret students' scores), based on that student's expected total raw score.

Mathematically, the TCC is the sum of all item/task characteristic functions for the items and tasks included on the test form (Lord, 1980). Thus, the TCC depends on the item/task characteristic functions (Lord, 1980). The shape of the TCC depends on several factors, including the number and the characteristics of the items/tasks, the item response theory model used, and the values of the item/task parameters. Consequently, there is no explicit formula for the TCC, and there are no parameters for the curve (Baker & Kim, 2017). As we present the Listening and Reading Online ACCESS tests in a multistage adaptive format and they are not fixed test forms, it is not appropriate to present TCCs for these tests.

Since raters use a polytomous scoring scale for Writing and Speaking tasks, the shapes of the TCCs for these tests are also affected by the parameter values for the individual categories on the scoring tools that raters use to evaluate students' responses to the tasks. These scoring tools have more score categories than the scoring schemes used for evaluating students' responses to multiple-choice items, which we typically score using just two categories— "right" or "wrong." By contrast, the Writing and Speaking rating scales have multiple score categories. For Writing, the rating scale has six whole score categories with an additional three in-between "plus" score categories, for a total of nine possible score points; for Speaking, the rating scale has five score categories. Therefore, the student ability measures for the Writing and Speaking domains will span a wide logit range (e.g., for the grade 1 Tier A Writing test, the student ability measures shown on the horizontal axis of Figure 2.8.3.1.1 range from -6 logits to 7 logits, a 13-logit spread).

Ideally, a TCC will be a smooth monotonically, or continuously increasing, S-shaped probability curve. However, when raters use multicategory rating scales to evaluate students' responses, they frequently do not assign equal numbers of scores in each of those categories. Consequently, the resulting adjacent score category boundaries may not be equidistant, and, indeed, in some cases, they may even be far apart if raters assign few scores in certain categories. In this situation, the curve of the TCC is likely to be somewhat bumpy or uneven across the student ability continuum. (The closer the adjacent score category boundaries are, the smoother the rise of the TCC along the student ability continuum.) Additionally, for some tests, the TCC may rise in a smooth S-shaped curve over the initial segment of the student ability continuum, but then plateau in the area between the boundaries of adjacent score categories before rising smoothly again, which would reflect the raters' uneven use of the score categories on the rating scale. We see this pattern in the TCCs for the Writing and Speaking

tests. The TCCs for other tests that include open-ended tasks, such as the National Assessment of Educational Progress Writing assessment (Muraki, 1993), often have this shape.

There are five vertical lines in each of the TCC figures indicating, for each test form, the cut scores for the highest grade in each grade-level cluster, dividing each figure into six sections that denote the WIDA proficiency levels (PL 1–PL 6) for the domain. As would be expected, higher raw scores are required for placement in higher proficiency levels. The relative width of each section between the cut score lines gives an indication of how many raw score points a student must achieve to be placed into a WIDA proficiency level.

2.8.1 Listening

The ACCESS 2.0 Online Listening test is a multistage adaptive assessment. As students do not all take the same set of items in the test, no test characteristic curve is presented.

2.8.2 Reading

The ACCESS 2.0 Online Reading test is a multistage adaptive assessment. As students do not all take the same set of items in the test, no test characteristic curve is presented.

2.8.3 Writing

2.8.3.1 Grade 1

Figure 2.8.3.1.1

Test Characteristic Curve: Writ 1 A S602 Online

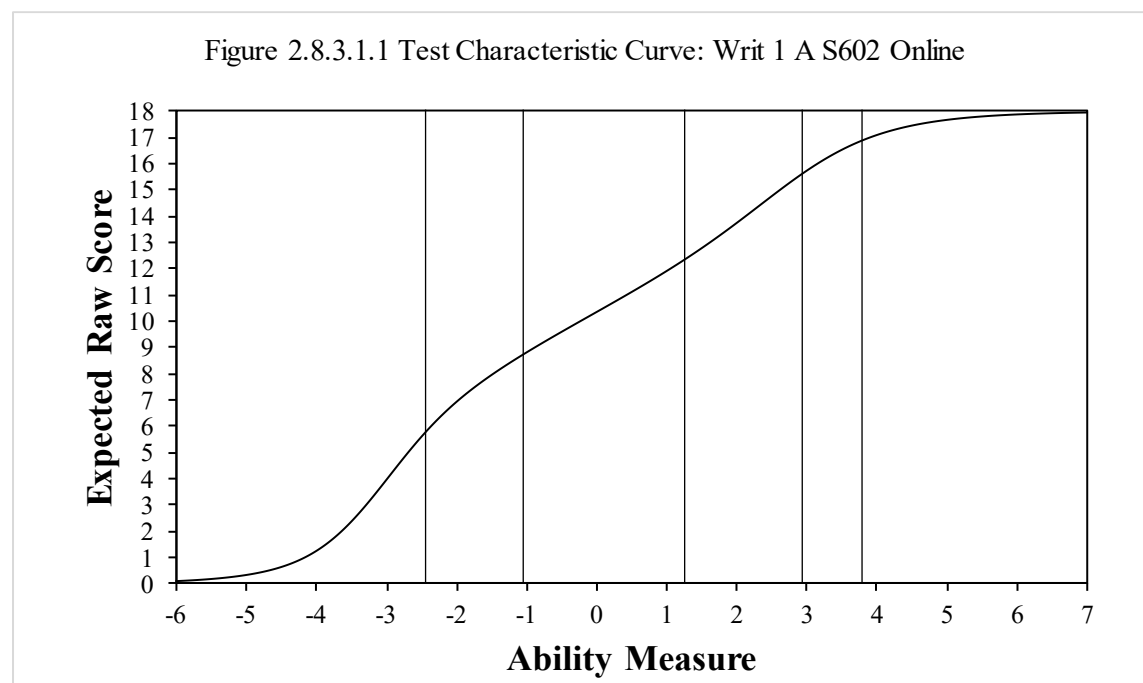
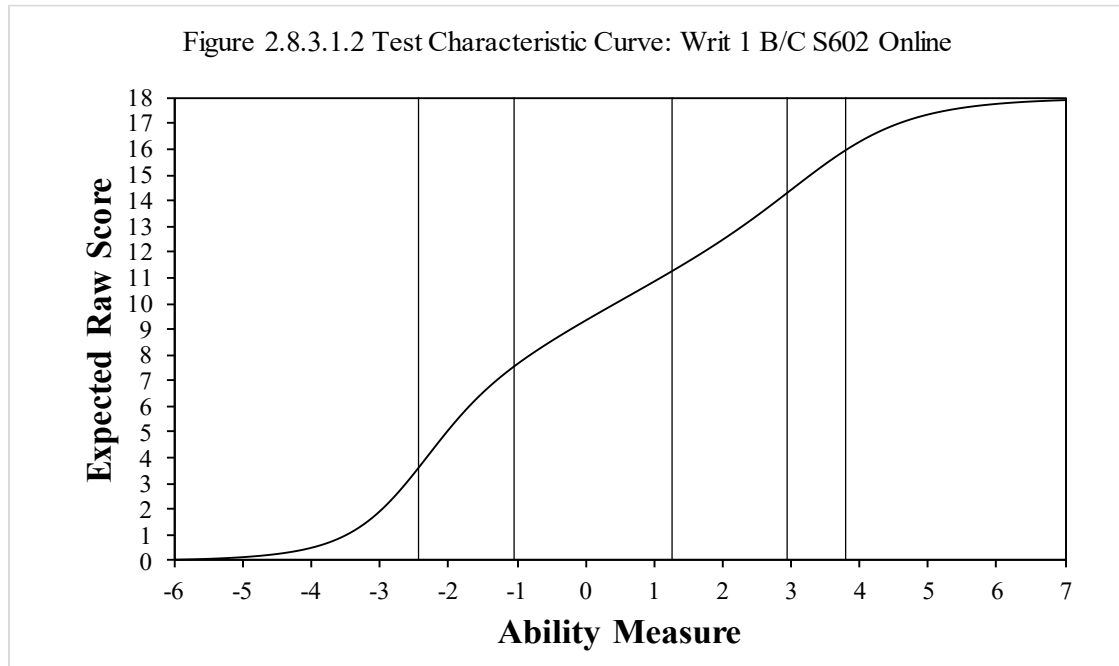


Figure 2.8.3.1.2

Test Characteristic Curve: Writ 1 B/C S602 Online



2.8.3.2 Grades 2–3

Figure 2.8.3.2.1

Test Characteristic Curve: Writ 2–3 A S602 Online

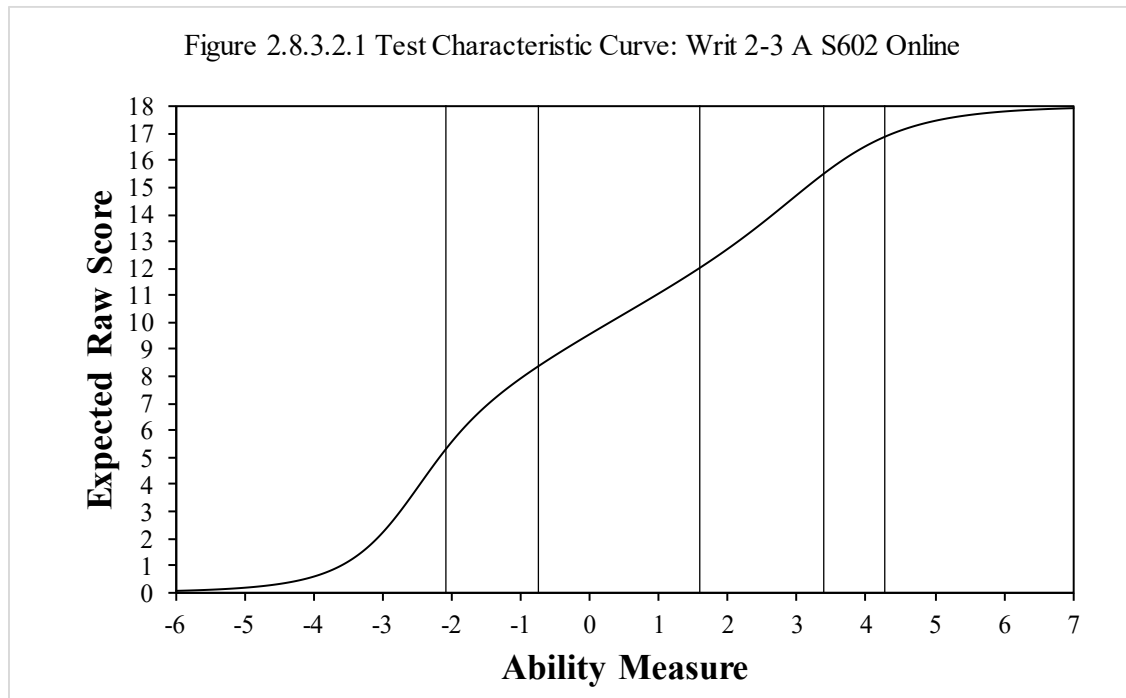
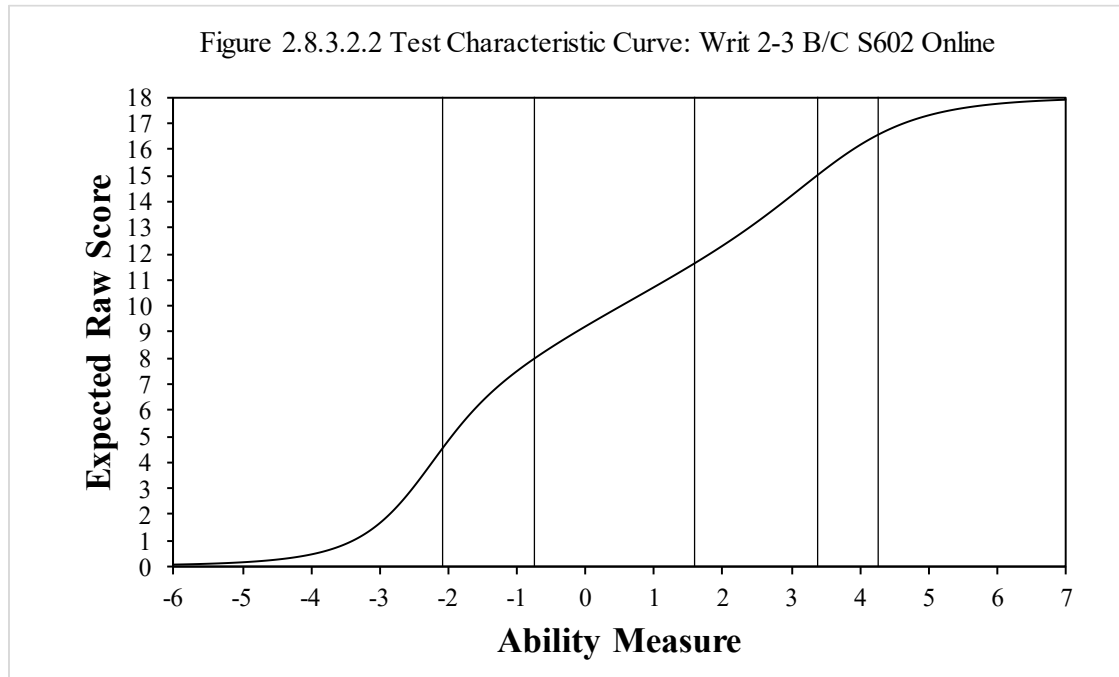


Figure 2.8.3.2.2.

Test Characteristic Curve: Writ 2–3 B/C S602 Online



2.8.3.3 Grades 4–5

Figure 2.8.3.3.1

Test Characteristic Curve: Writ 4–5 A S602 Online

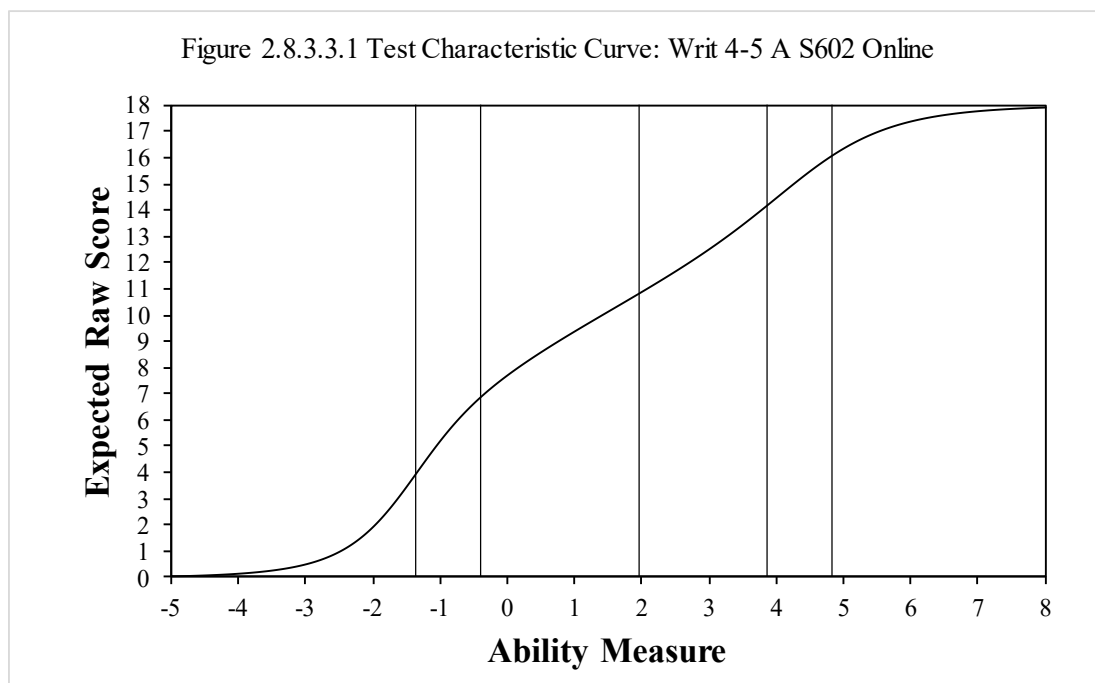
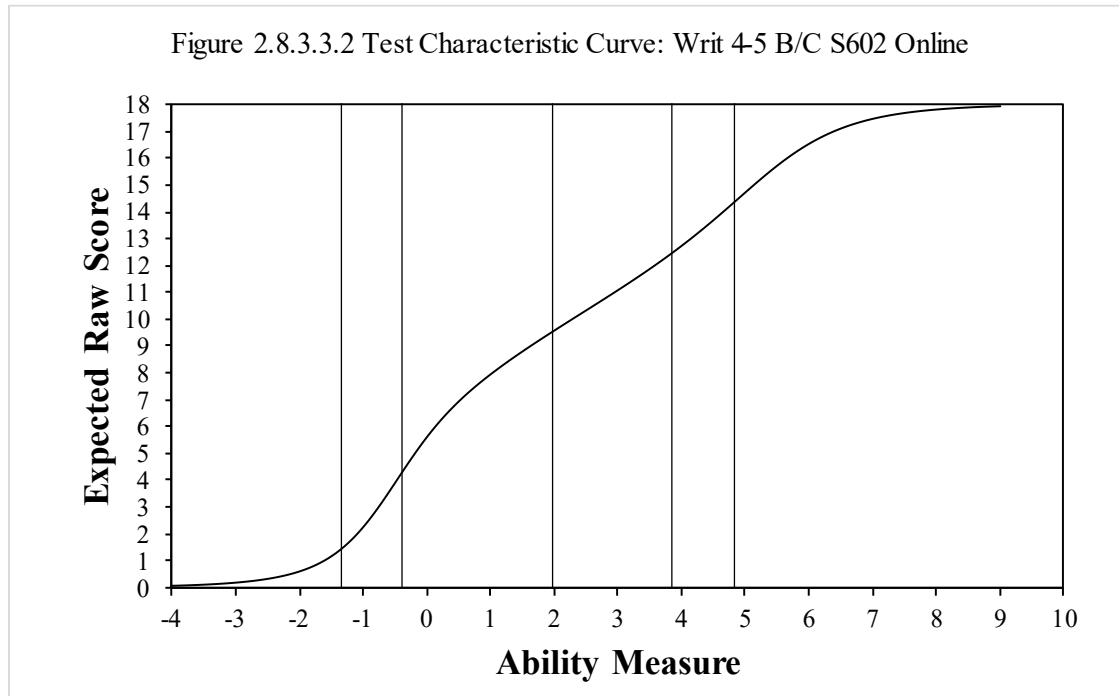


Figure 2.8.3.3.2

Test Characteristic Curve: Writ 4–5 B/C S602 Online



2.8.3.4 Grades 6–8

Figure 2.8.3.4.1

Test Characteristic Curve: Writ 6–8 A S602 Online

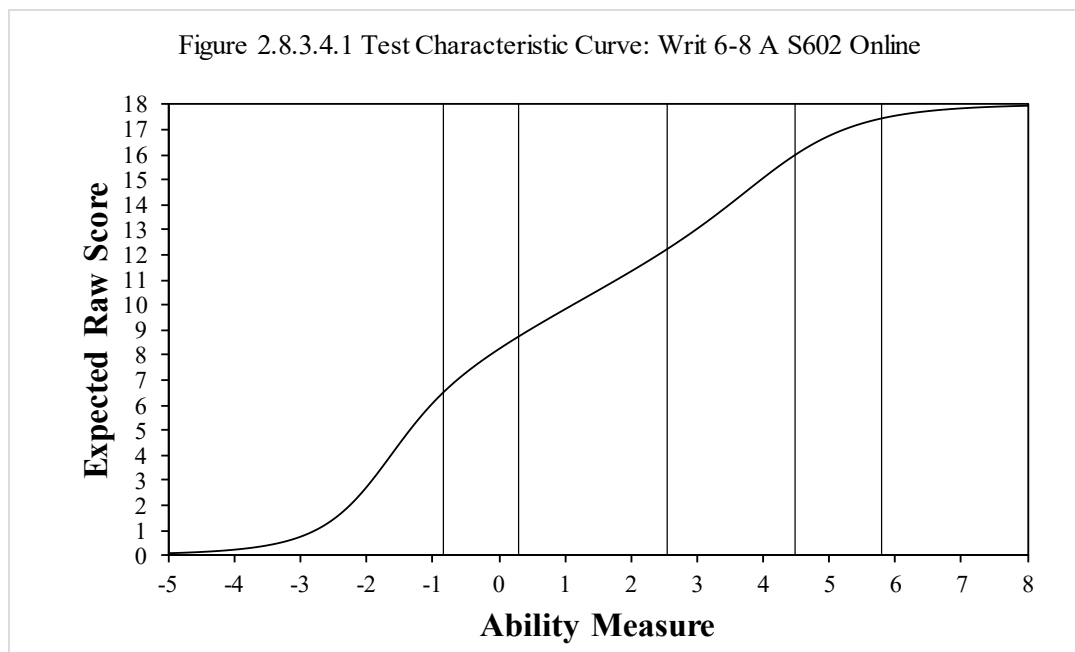
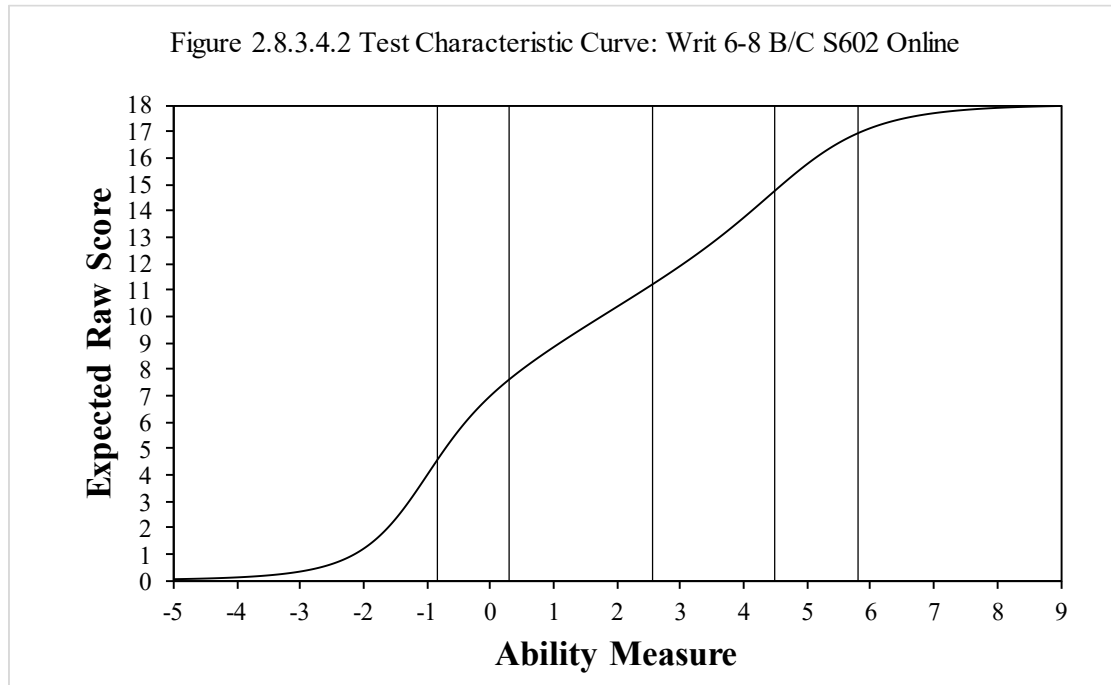


Figure 2.8.3.4.2

Test Characteristic Curve: Writ 6–8 B/C S602 Online



2.8.3.5 Grades 9–12

Figure 2.8.3.5.1

Test Characteristic Curve: Writ 9–12 A S602 Online

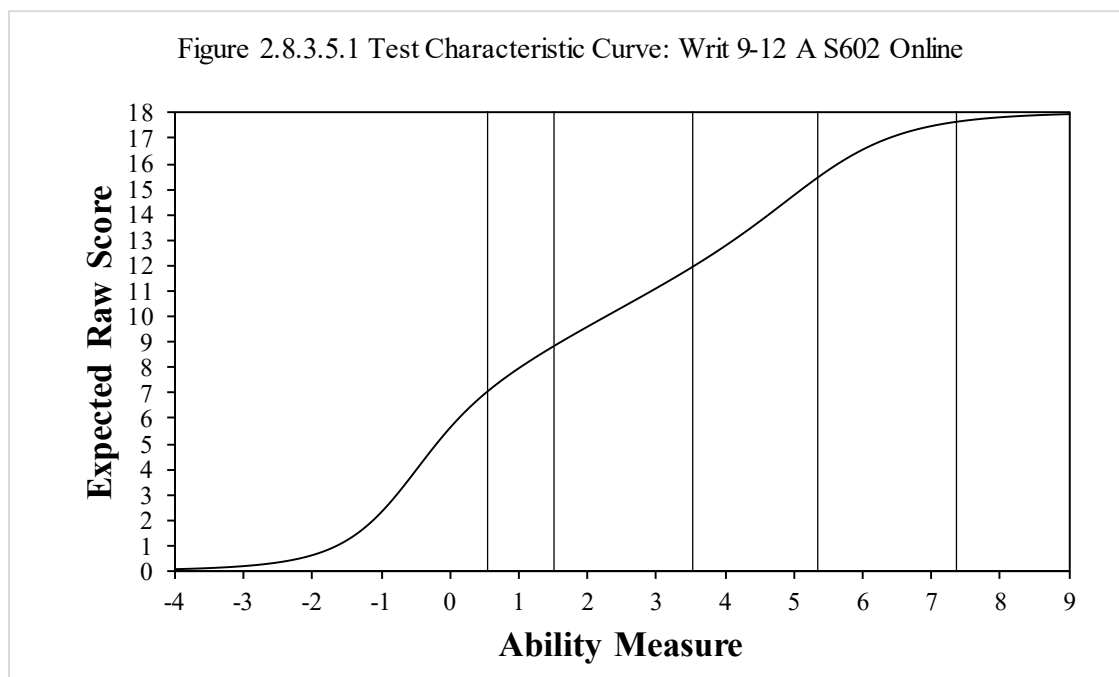
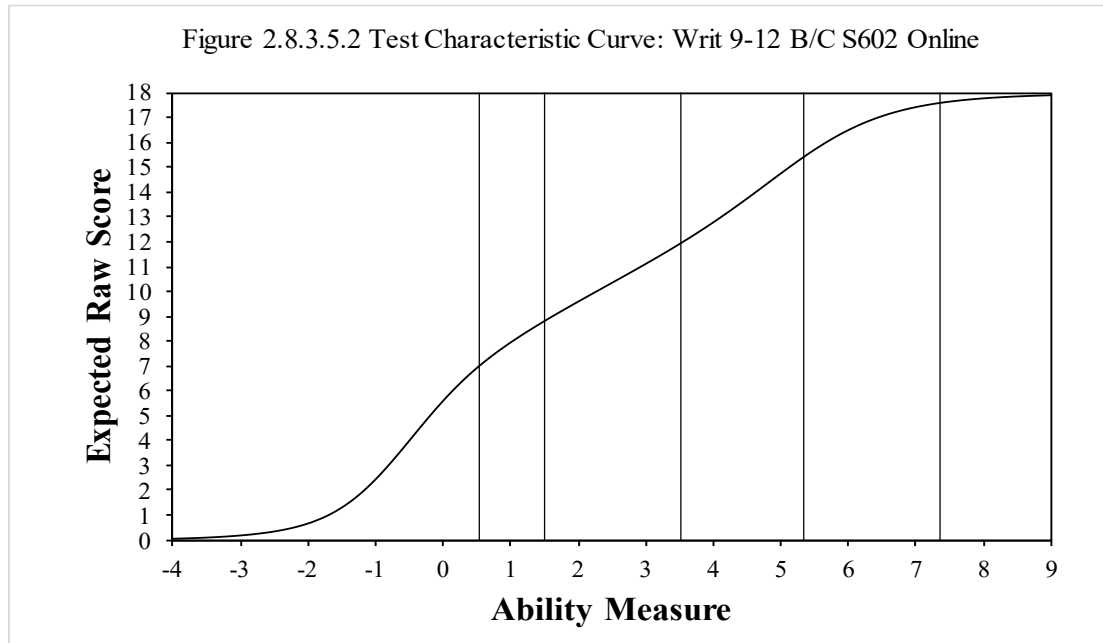


Figure 2.8.3.5.2

Test Characteristic Curve: Writ 9–12 B/C S602 Online



2.8.4 Speaking

2.8.4.1 Grade 1

Figure 2.8.4.1.1

Test Characteristic Curve: Spek 1 Pre-A S602 Online

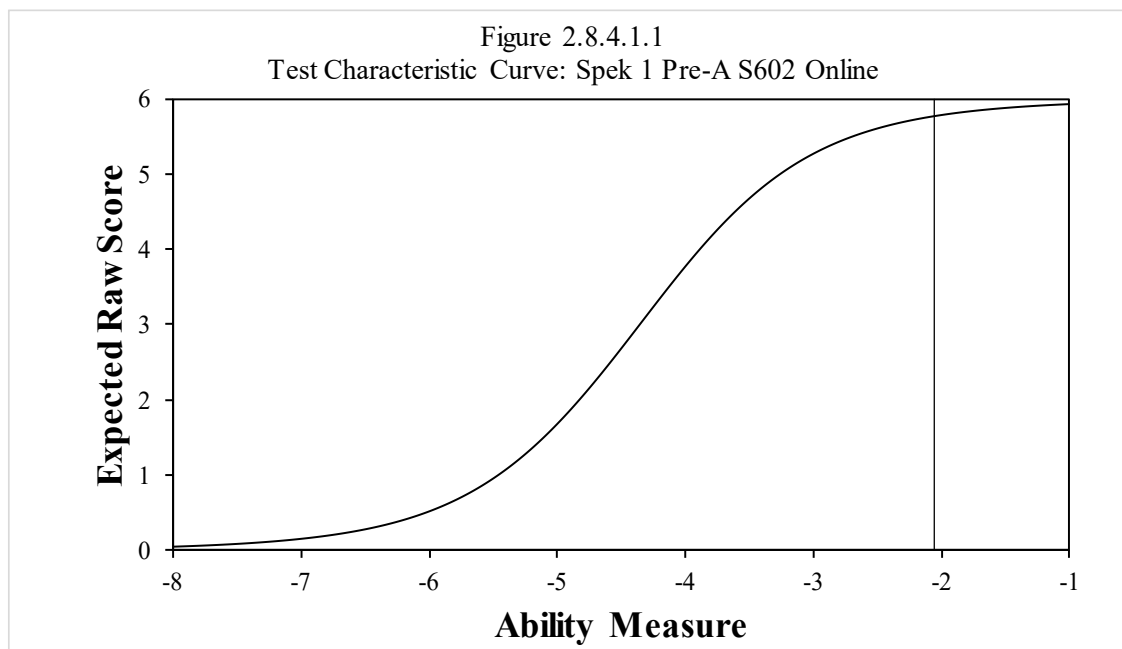


Figure 2.8.4.1.2

Test Characteristic Curve: Spek 1 A S602 Online

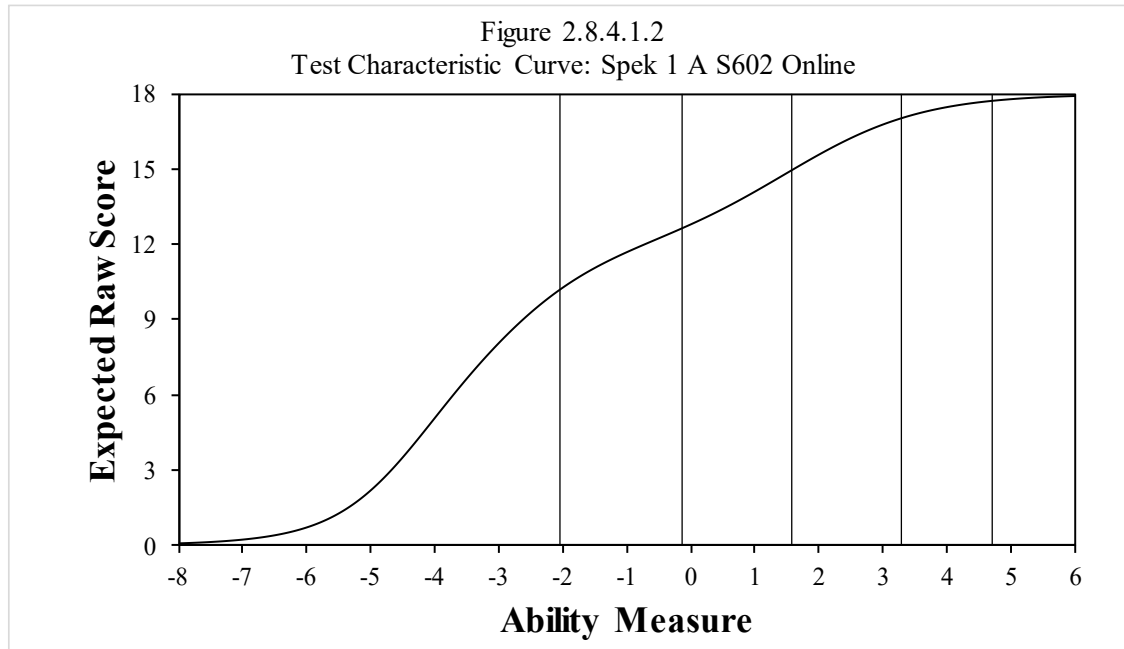
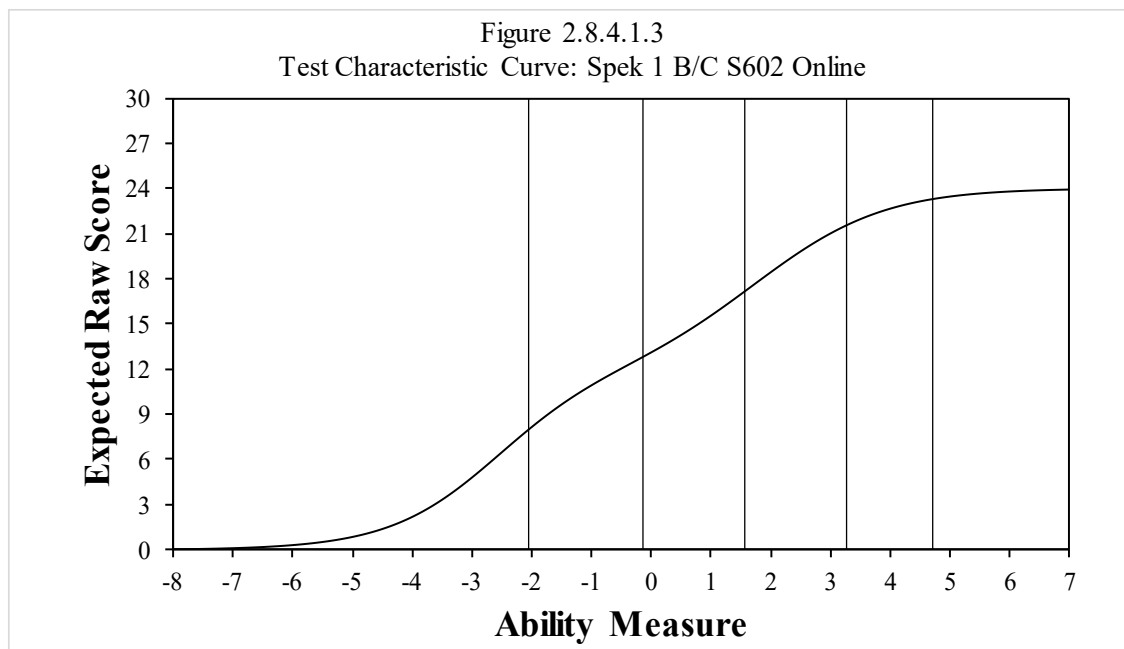


Figure 2.8.4.1.3

Test Characteristic Curve: Spek 1 B/C S602 Online



2.8.4.2 Grades 2–3

Figure 2.8.4.2.1

Test Characteristic Curve: Spek 2–3 Pre-A S602 Online

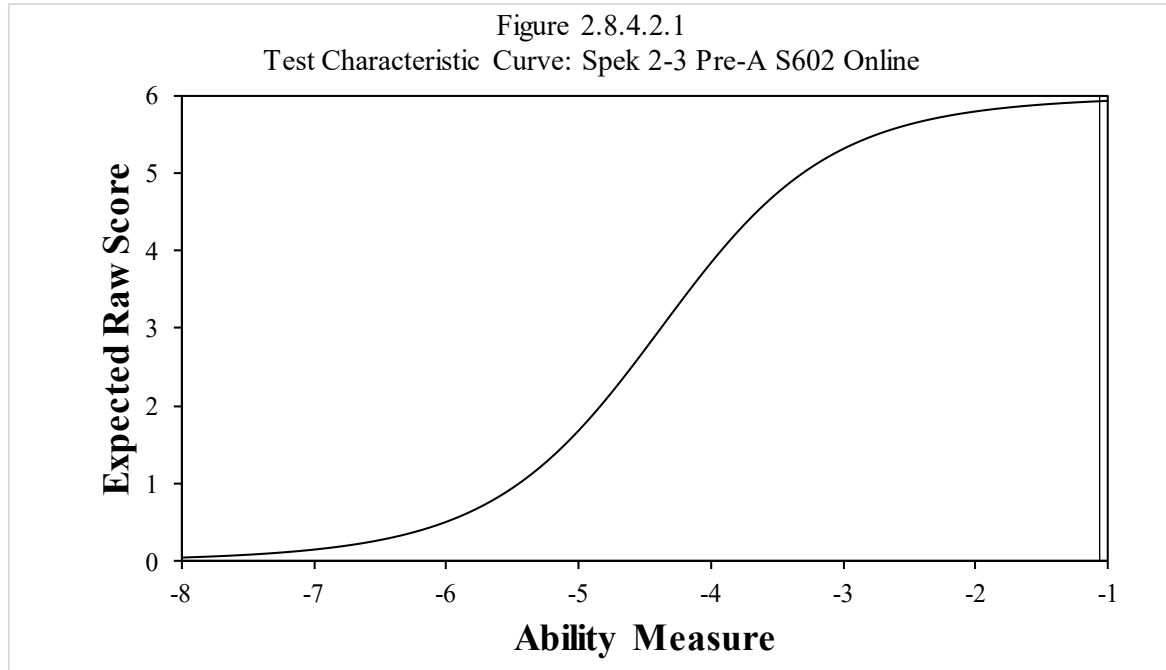


Figure 2.8.4.2.2

Test Characteristic Curve: Spek 2–3 A S602 Online

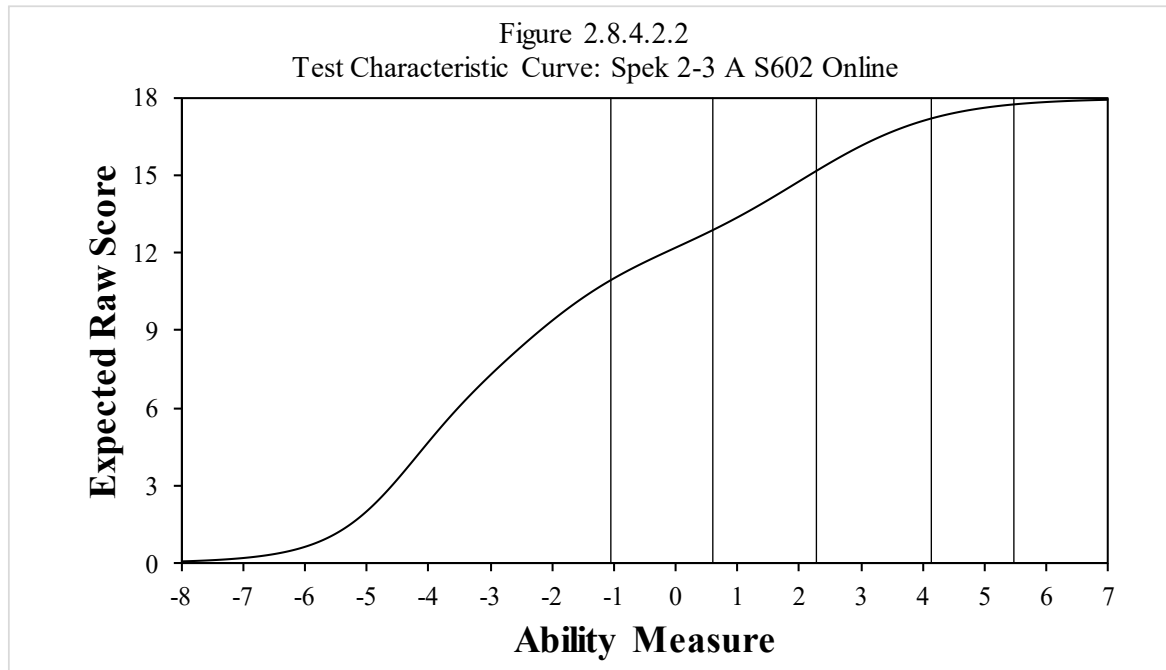
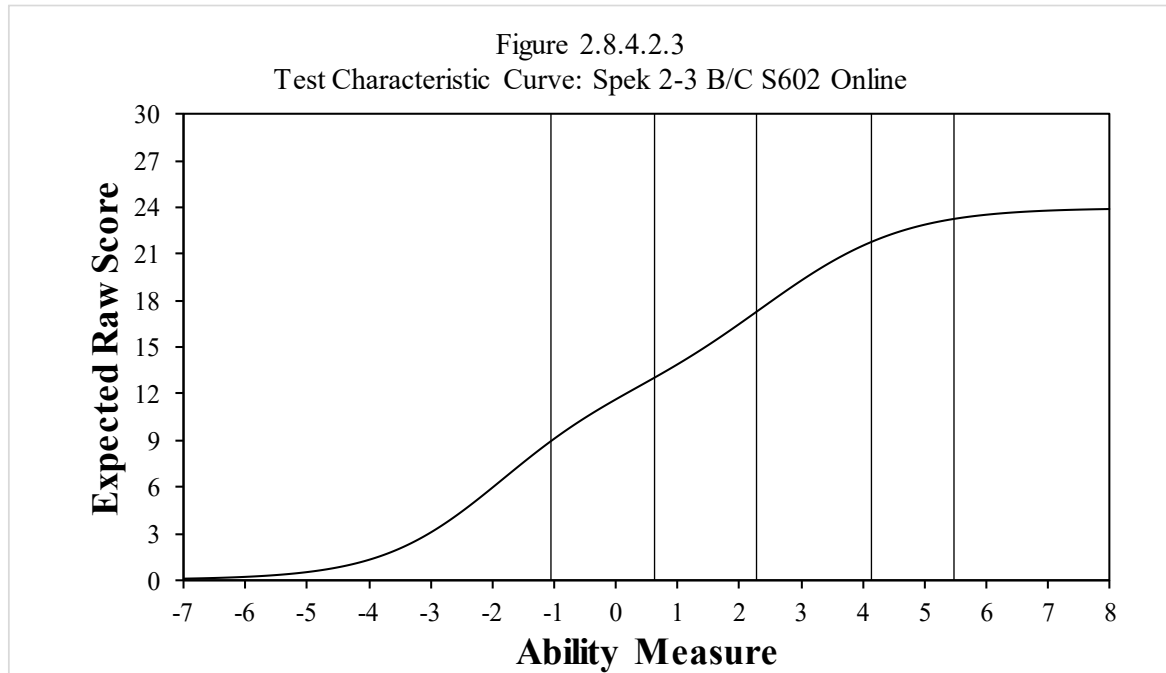


Figure 2.8.4.2.3

Test Characteristic Curve: Spek 2-3 B/C S602 Online



2.8.4.3 Grades 4-5

Figure 2.8.4.3.1

Test Characteristic Curve: Spek 4-5 Pre-A S602 Online

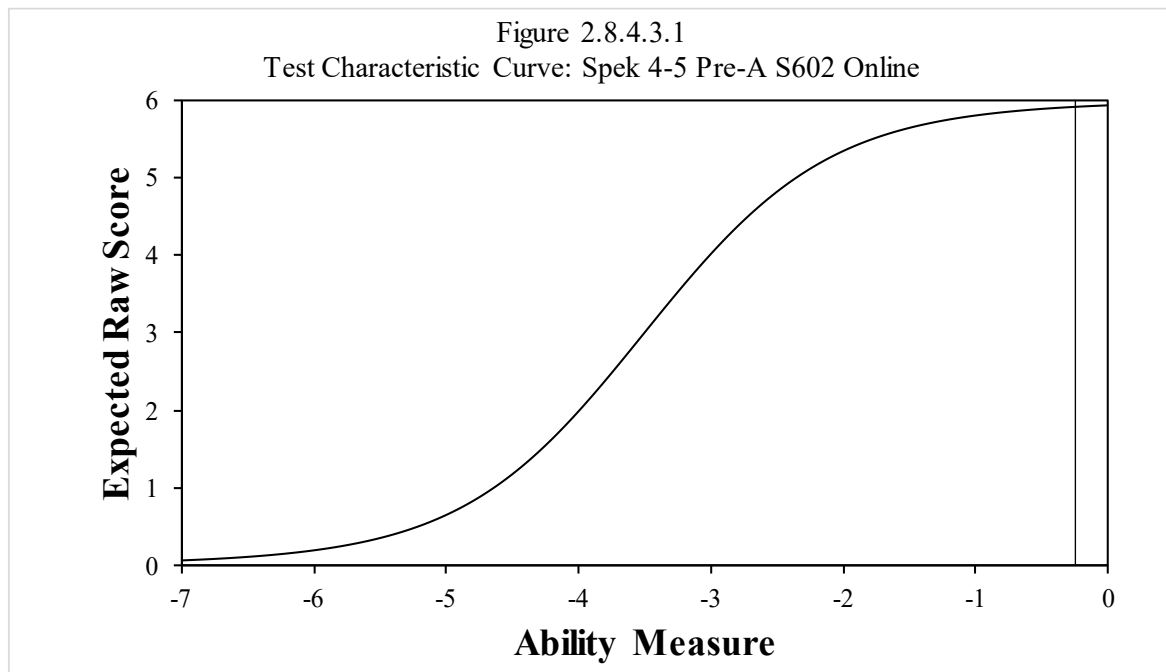


Figure 2.8.4.3.2

Test Characteristic Curve: Spek 4-5 A S602 Online

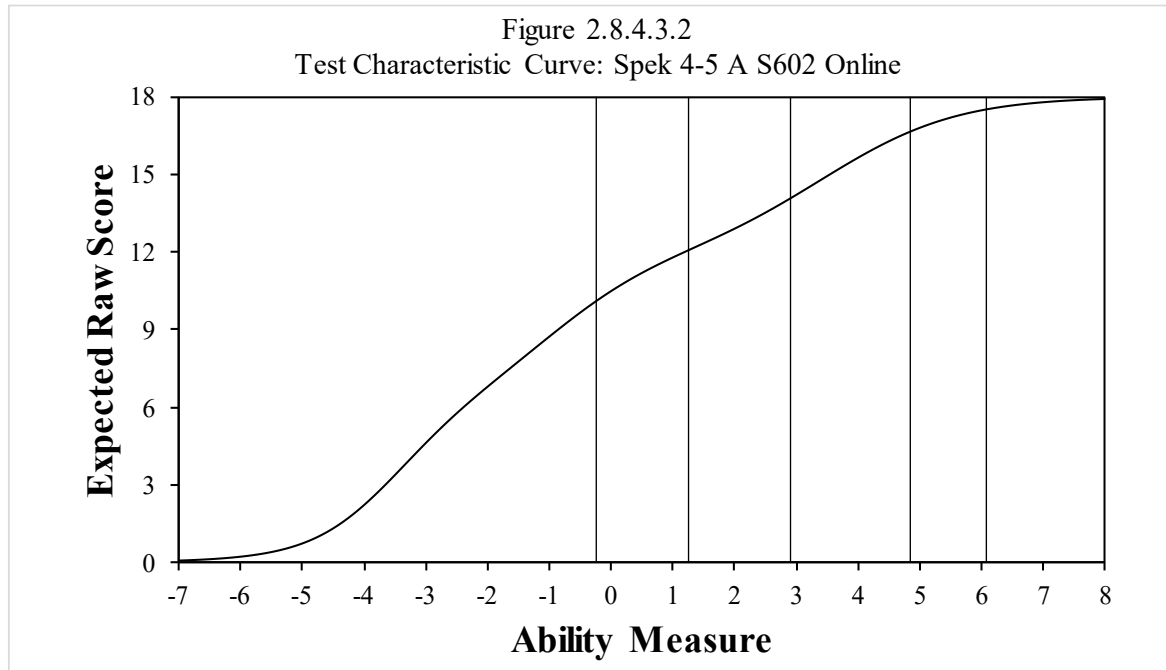
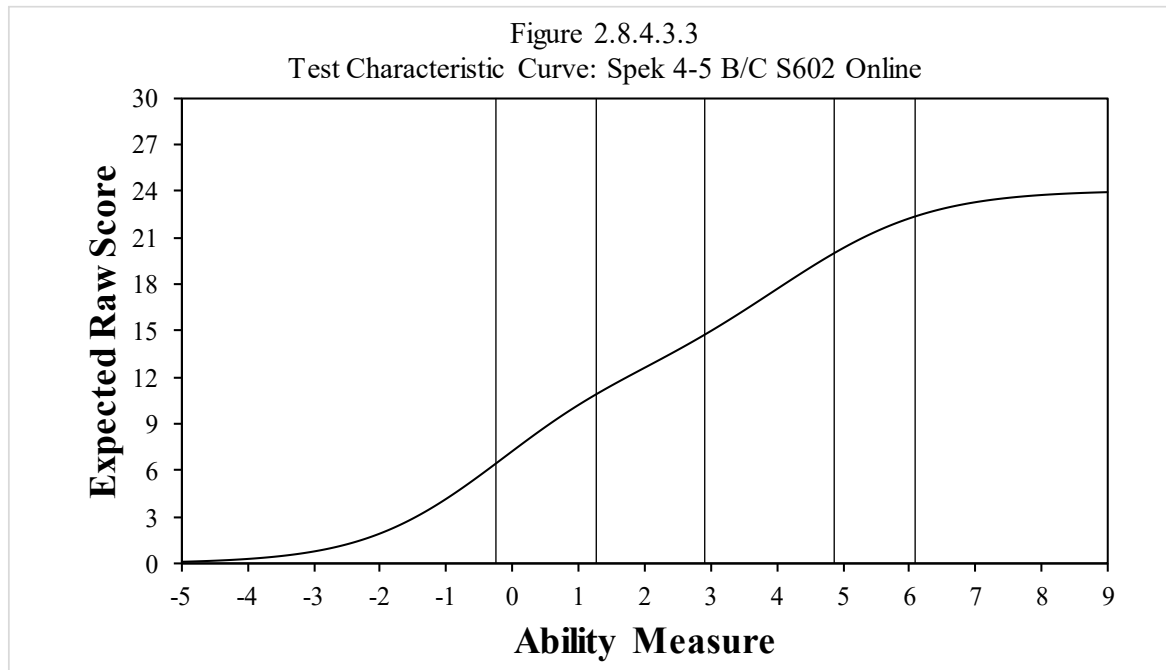


Figure 2.8.4.3.3

Test Characteristic Curve: Spek 4-5 B/C S602 Online



2.8.4.4 Grades 6–8

Figure 2.8.4.4.1

Test Characteristic Curve: Spek 6–8 Pre-A S602 Online

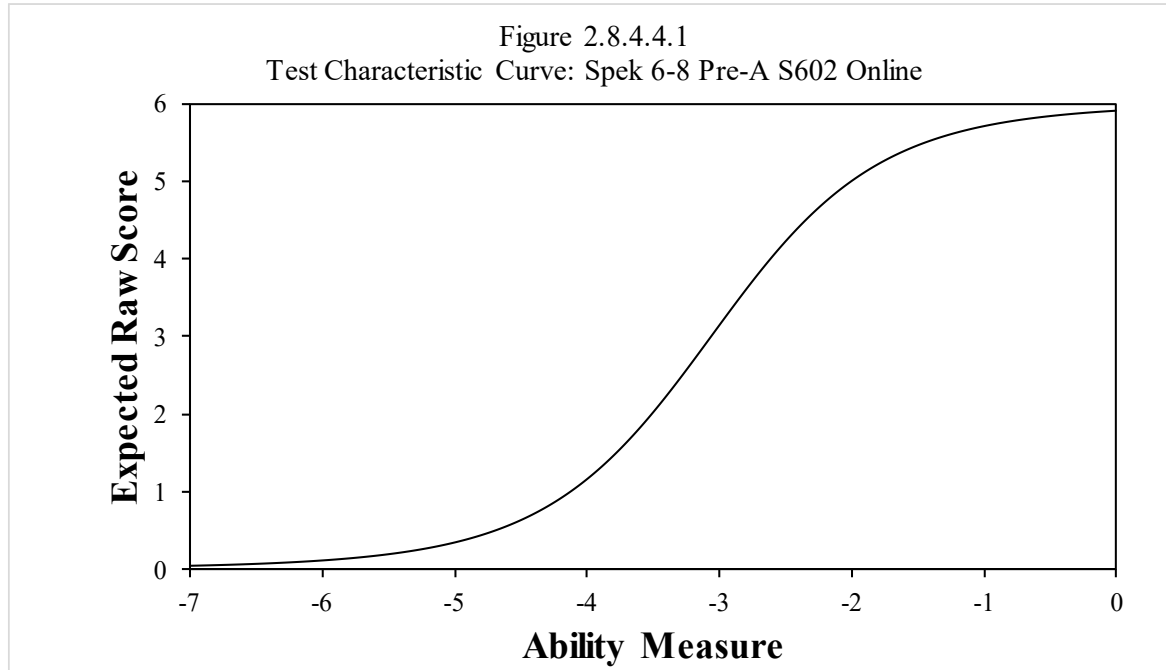


Figure 2.8.4.4.2

Test Characteristic Curve: Spek 6–8 A S602 Online

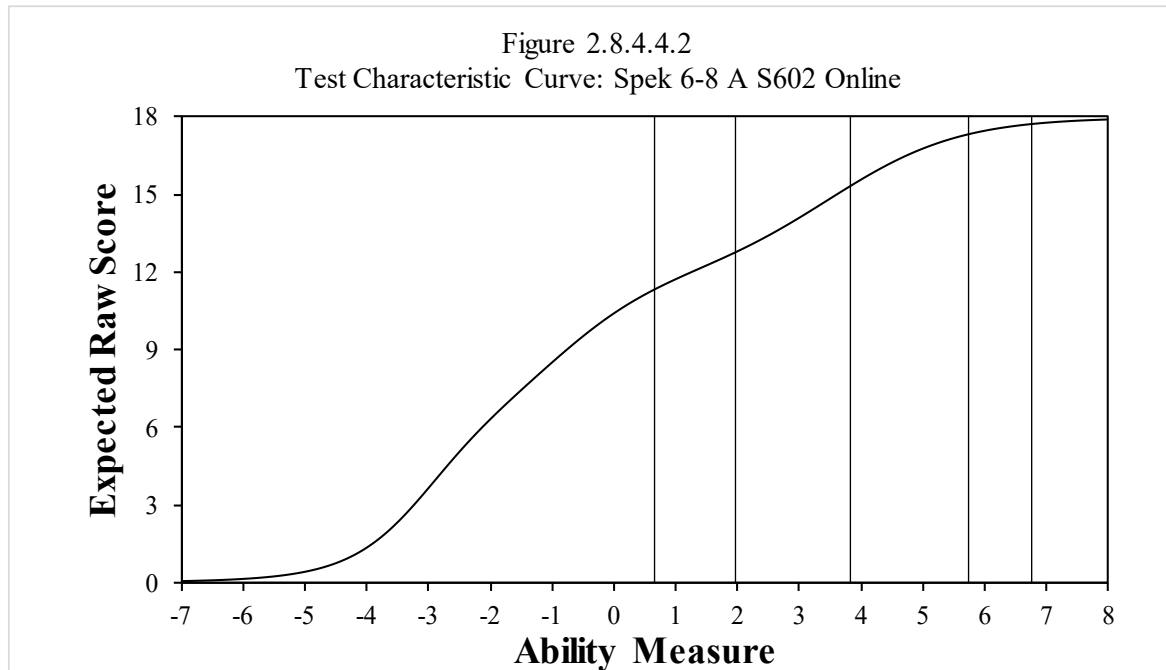
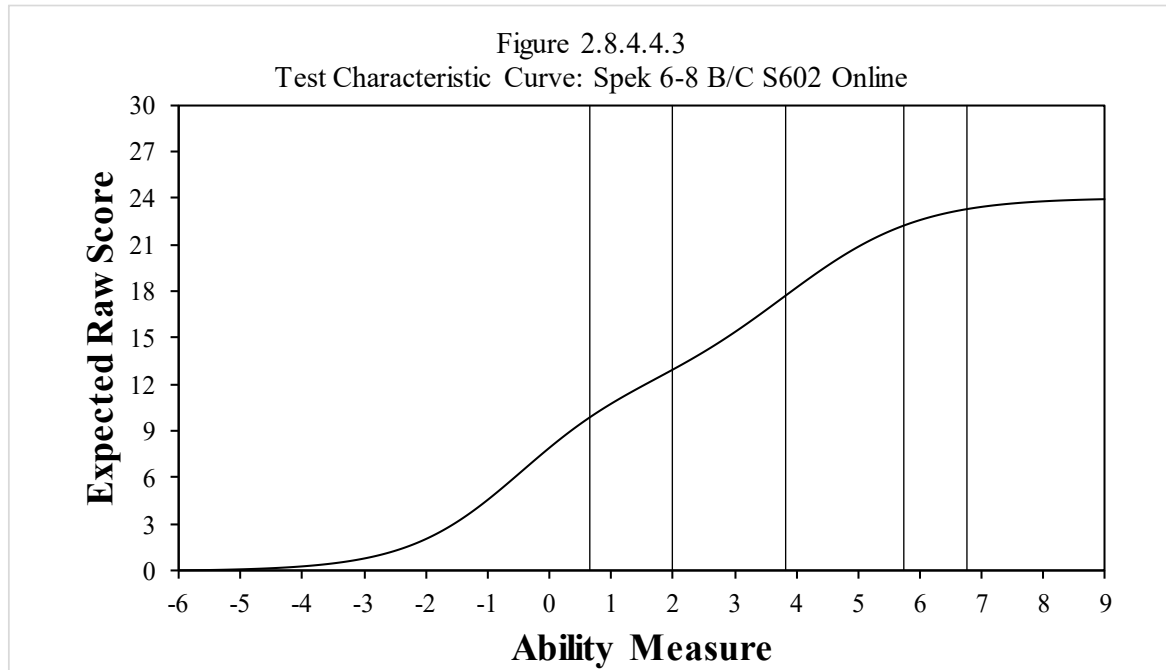


Figure 2.8.4.4.3

Test Characteristic Curve: Spek 6–8 B/C S602 Online



2.8.4.5 Grades 9–12

Figure 2.8.4.5.1

Test Characteristic Curve: Spek 9–12 Pre-A S602 Online

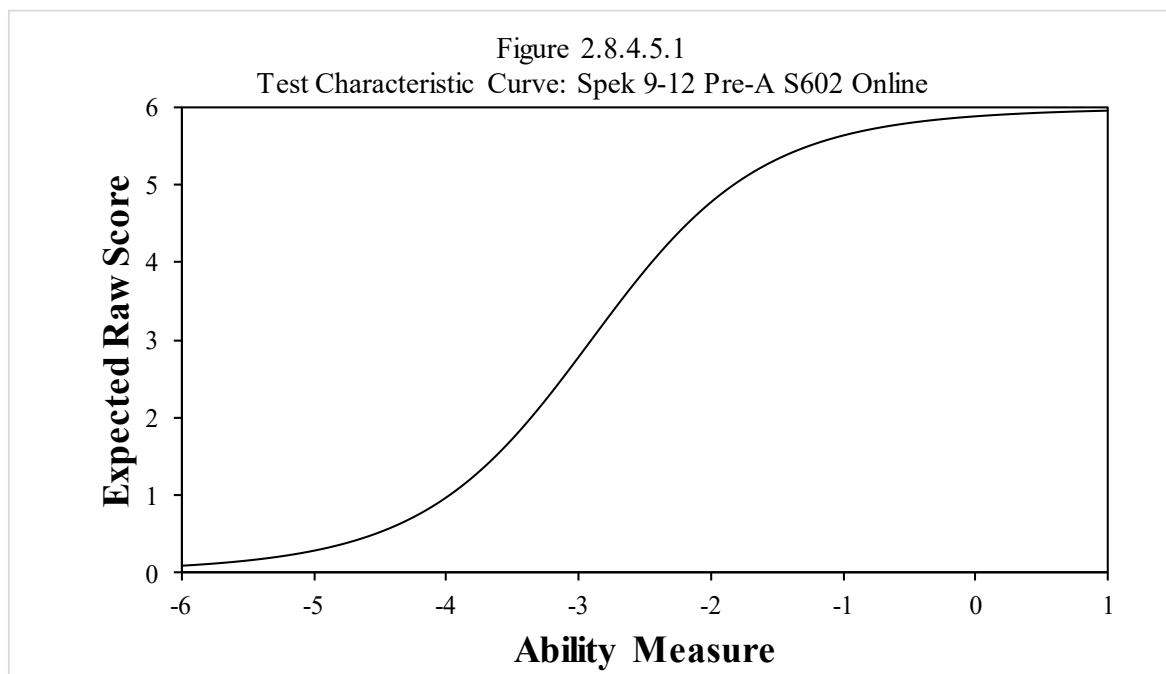


Figure 2.8.4.5.2

Test Characteristic Curve: Spek 9–12 A S602 Online

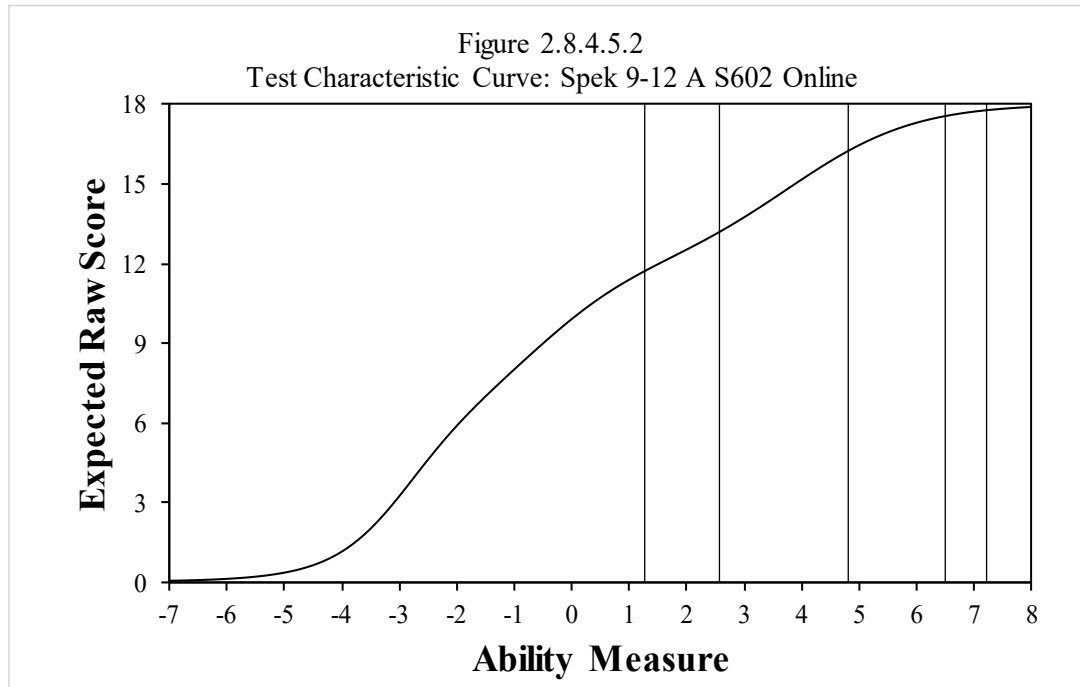
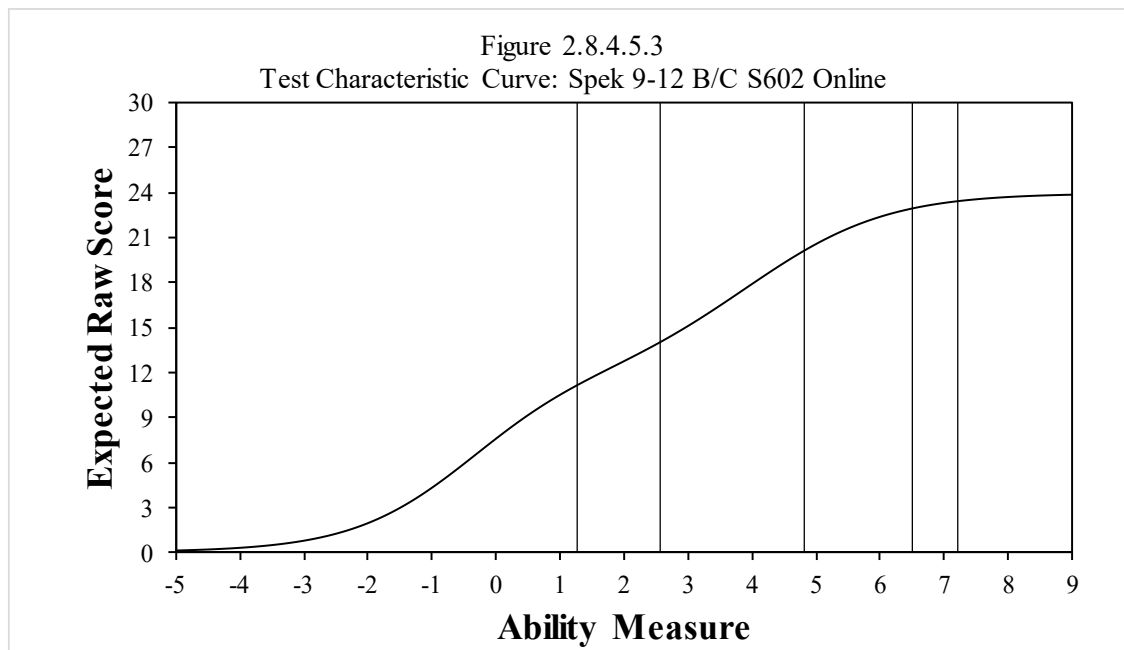


Figure 2.8.4.5.3

Test Characteristic Curve: Spek 9–12 B/C S602 Online



2.9 Test Information Function

With Rasch measurement models, as with any measurement model that is based on item response theory, one can use the item/task information function (Lord, 1980) to model the relationship between a student ability measure (in logits) and the amount of information that the students' responses to that item (or task) provides about that student's true ability. Tests perform differently for students who have differing levels of ability. Difficult items (or tasks) provide useful information for differentiating among higher-ability students but are not useful for differentiating among lower-ability students. Conversely, easy items (or tasks) provide useful information for differentiating among lower-ability students but not for differentiating among higher-ability students. Consequently, an item (or task) will provide maximum information when it is well targeted to the ability measure of the student (Reise, 1999).

The **item/task information function** indicates the amount of information that students' responses to that item (or task) provides to help reduce our uncertainty regarding a student's true ability measure. The more information we have about the ability measure, the more certain or confident we can be in that estimate of the student's ability. If the amount of information is large, that means that we have estimated with a higher degree of certainty a student whose true ability is at that level. Therefore, the ability measures for students whose scores lie within that region of the ability continuum will be reasonably close to their true values. Conversely, if the amount of information is small, that means that we have estimated with a lower degree of certainty the student whose true ability is at that level. Consequently, the ability measures for students whose scores lie within that region of the ability continuum will be further away from their true values.

Mathematically, for an item (or task), the amount of information for a given ability level is the reciprocal of the variance of the ability measure at the level. In other words, for that item (or task), the information value is the inverse squared of the standard errors of measurement for a given ability measure. Therefore, for that item (or task), the information value also provides information about the precision of the ability measure along the ability continuum.

The **test information function** (TIF) aggregates the item/task information functions across all the items (and/or tasks) on the test form or in the item pool. Since for an item (or task) the information value is the inverse squared of an ability measure's standard error of measurement, the TIF reflects, for the whole test, the standard error of measurement for all ability measures. When the TIF is presented graphically as the test information curve, it shows how well the test is measuring across the continuum of student ability in terms of the amount of information (i.e., certainty), or the amount of measurement precision, the test provides at each ability level. The higher the curve in a particular region of the ability continuum, the more information the test provides at the ability level.

Since the TIF is the sum of all item/task information functions on the test form (Lord, 1980), the TIF depends on the information functions (Lord, 1980) of the individual items/tasks included on the test form or in the item pool. The shape of the test information curve depends on several factors, including the number and characteristics of items/tasks, the item response theory model used, and the values of the item/task parameters. With some exceptions, there is

a general pattern to the shape of test information curves. Test information curves peak in the region of the student ability continuum where the test provides higher discrimination and more precise measurement as compared to other regions where the curve is less peaked, normally at the lower and upper ends of the ability continuum. When the test form consists of multiple-choice items such as in the Listening and Reading domains, the test information curve is usually unimodal.

The parameter values for the individual categories on the scoring tools that raters use to evaluate students' responses to the tasks, in addition to the factors mentioned earlier, affect the shape of the test information curves for the Writing and Speaking tests. Accordingly, some refer to these test information curves as "category information functions" (Engelhard & Wind, 2018). The scoring scales that the raters use have more score categories than the scoring schemes used for evaluating students' responses to multiple-choice items, which typically have just two categories— "right" or "wrong." Additionally, we designed the scoring scales to measure a wide range of student performance on a task. Consequently, the resulting adjacent score category boundaries may not be equidistant, and, indeed, in some cases, they may even be far apart if raters assign few scores in certain categories. In this situation, a test information curve will have one (or more) dips in the region(s) between the adjacent score category boundaries, indicating the loss of information in the corresponding ability range(s) and a decrease in the amount of information that certain score categories provide (Engelhard & Wind, 2018). Therefore, the shape of a test information curve for an ACCESS Writing or Speaking test may not be unimodal and instead may have two (or more) peaks. For example, suppose that a test information curve reveals a dip in the region of the student writing ability continuum where raters would have assigned a score of 3. That suggests that students who received a score of 3 may have displayed potentially substantively meaningful differences in writing ability that the raters were not able to adequately distinguish when they used the 9-point Writing Scoring Scale to assign scores or, alternatively, that the score categories did not describe salient characteristics of students' writing that would make it possible for the raters to distinguish reliably among the students' responses in that region of the student ability continuum (Engelhard & Wind, 2018, pp. 316–319). The ACCESS Writing and Speaking tests are not the only assessments that have test information curves with these unusual shapes. The test information curves for other tests composed of open-ended tasks, such as the National Assessment of Educational Progress Writing assessment, also show a similar "dipping" pattern (Muraki, 1993).

The figures in this section plot the TIFs and show graphically the amount of information that the test provided across the continuum of student ability. For each test form, the five vertical lines in the figure indicate the ACCESS cut scores for the highest grade in each grade-level cluster, dividing the figure into six sections denoting the WIDA proficiency levels (PL 1–PL 6) for the domain. The test information curve and the corresponding ACCESS cut-score lines are both expressed on the ACCESS logit scale. Note that for the Speaking test, in Tier Pre-A, all scores are within the PL 1.0 range, so for some graphs, no vertical lines are showing the cut scores between proficiency levels.

The inclusion of the ACCESS cut-score lines in these figures is meant only to facilitate the visual interpretation of the test information curves relative to the ACCESS cut scores by

domains. These lines provide a benchmark for WIDA and CAL assessment experts to examine the ability range for which each test seems to be more (or less) accurate in estimating student ability. Readers should note that most states that use ACCESS for ELLs do not make reclassification decisions based solely on students' domain scale scores. Rather, the majority of these states set their reclassification (or exit) criterion based on a student's Overall composite scale score, which is a weighted sum of a student's four domain scale scores. Only a few states use those four domain scale scores in addition to the student's Overall composite scale score when making a reclassification decision. Therefore, from the WIDA policy perspective, it is more important to ensure that we minimize the measurement error near the cut score that most states use to set their reclassification criterion on the Overall composite scale score. We report the conditional standard errors of measurement (CSEMs) for the Overall composite scale scores in Section 5.6.

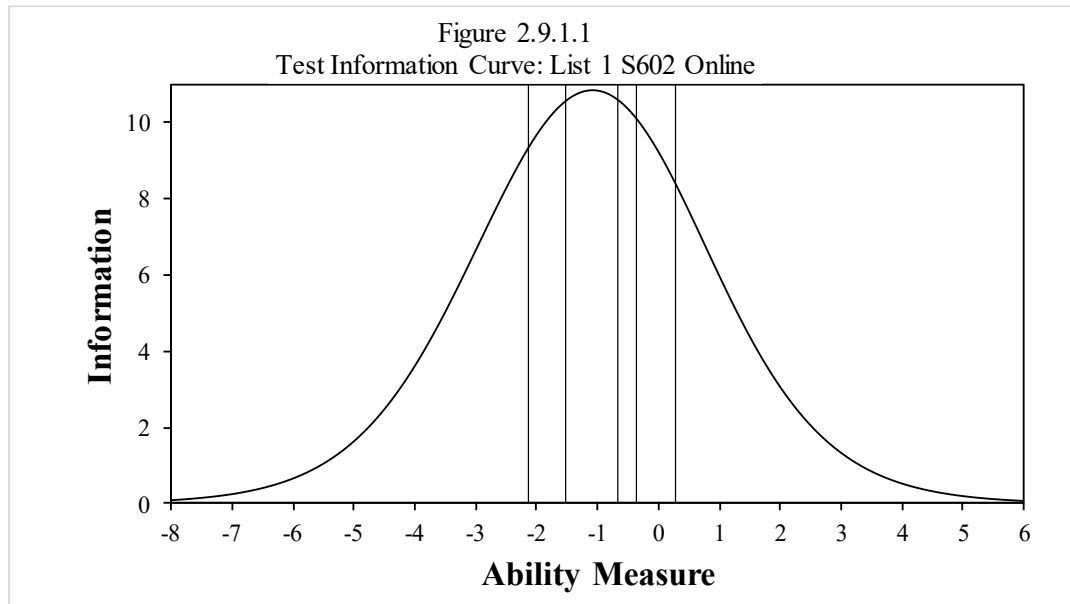
In addition to the TIF graphs by tier, for the Writing and Speaking tests, in the same graph we provide plots of the TIFs across tiers, by grade-level cluster. Test users may find it useful to compare the ability ranges across tiers, noting for each tier where the curve displays its highest peaks (i.e., where the most measurement information is provided). For example, as shown in Figure 2.9.3.1.3, the test information curve across tiers for Writing grade 1 reveals that the Writing grade 1 Tier A form provided more information about student ability measures that were either just below the PL 2 cut score or just below the PL 5 cut score. By contrast, the Writing grade 1 Tier B/C form provided more information about the student ability measures that were either just above the PL 2 cut score or just above the PL 5 cut score. The plot also shows that the Writing grade 1 Tier A form provided more information for those student ability measures in the lowest range (i.e., ability measures of -0.5 logits or lower), while the Writing grade 1 Tier B/C form provided more information than the grade 1 Tier A form for the rest of the student ability measures, especially those in the higher ability range. Lastly, consistent with the purposes of the test design, there is also considerable overlap between the ranges of writing ability that the two forms cover.

2.9.1 Listening

2.9.1.1 Grade 1

Figure 2.9.1.1

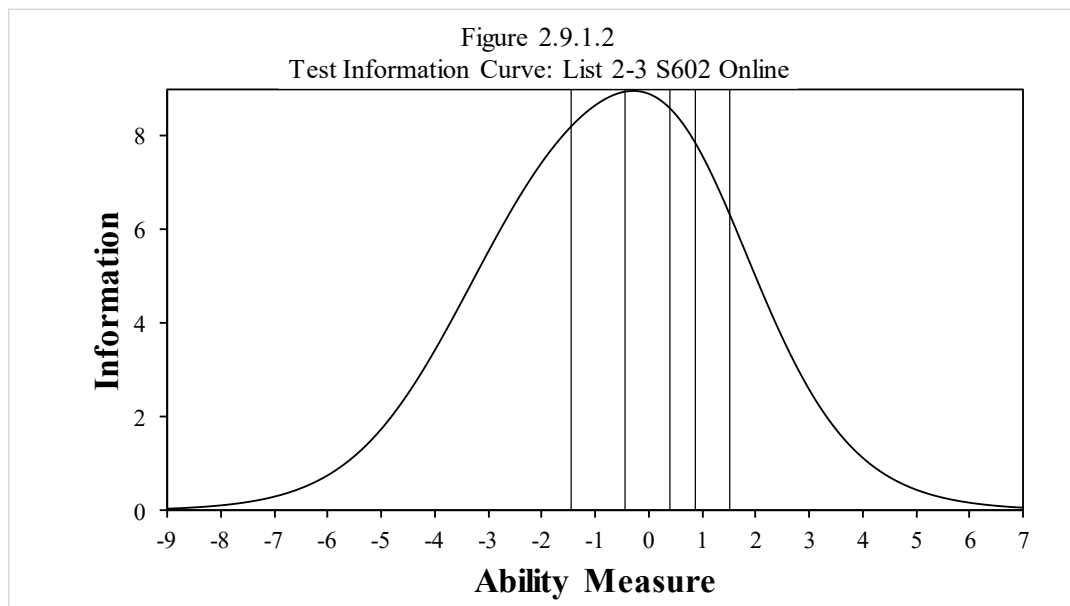
Test Information Curve: List 1 S602 Online



2.9.1.2 Grades 2–3

Figure 2.9.1.2

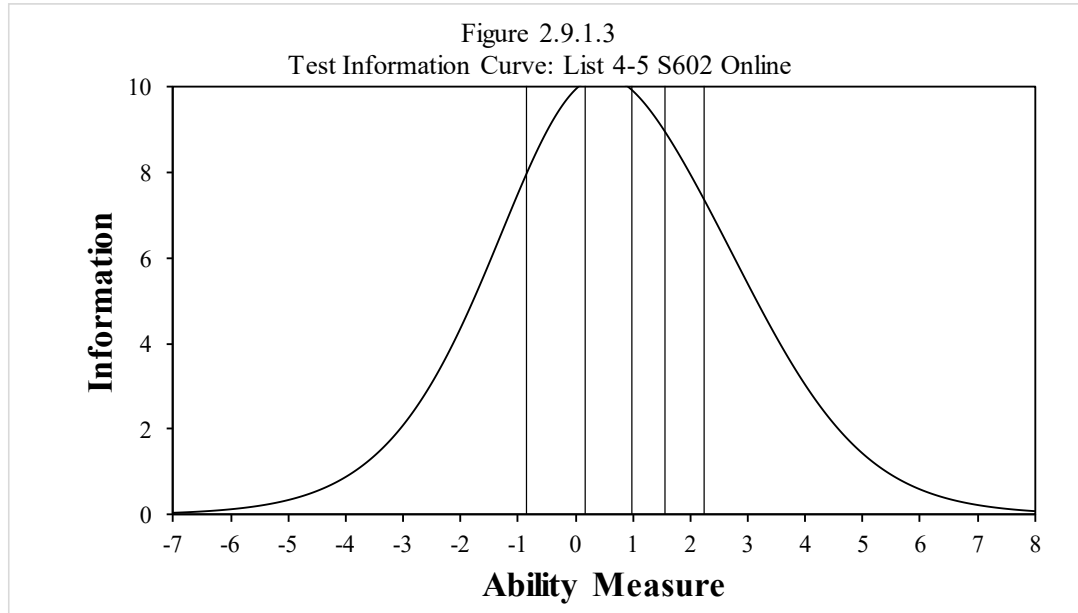
Test Information Curve: List 2–3 S602 Online



2.9.1.3 Grades 4–5

Figure 2.9.1.3

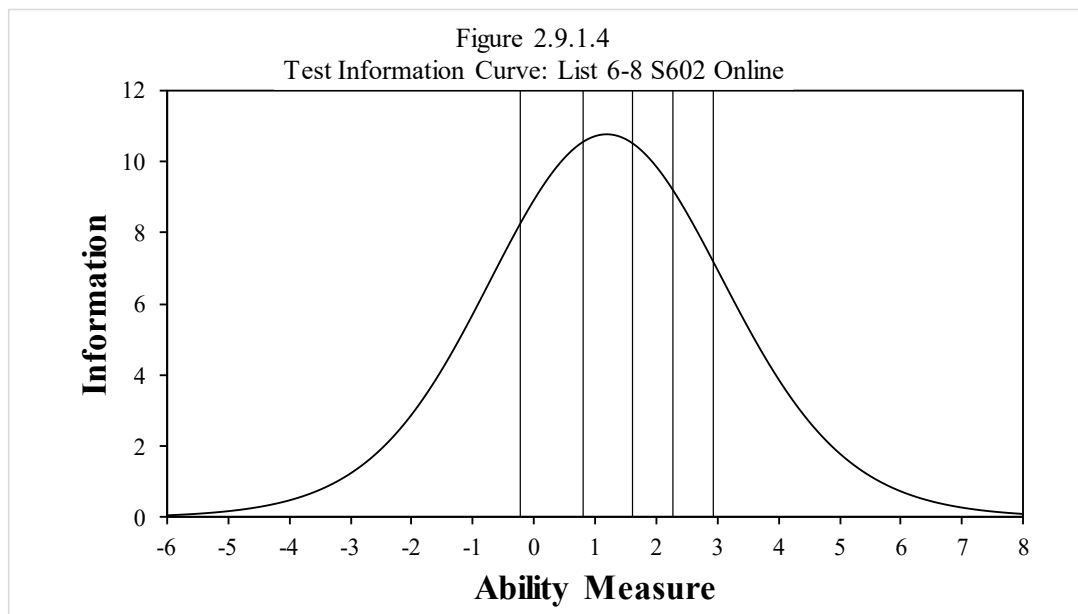
Test Information Curve: List 4–5 S602 Online



2.9.1.4 Grades 6–8

Figure 2.9.1.4

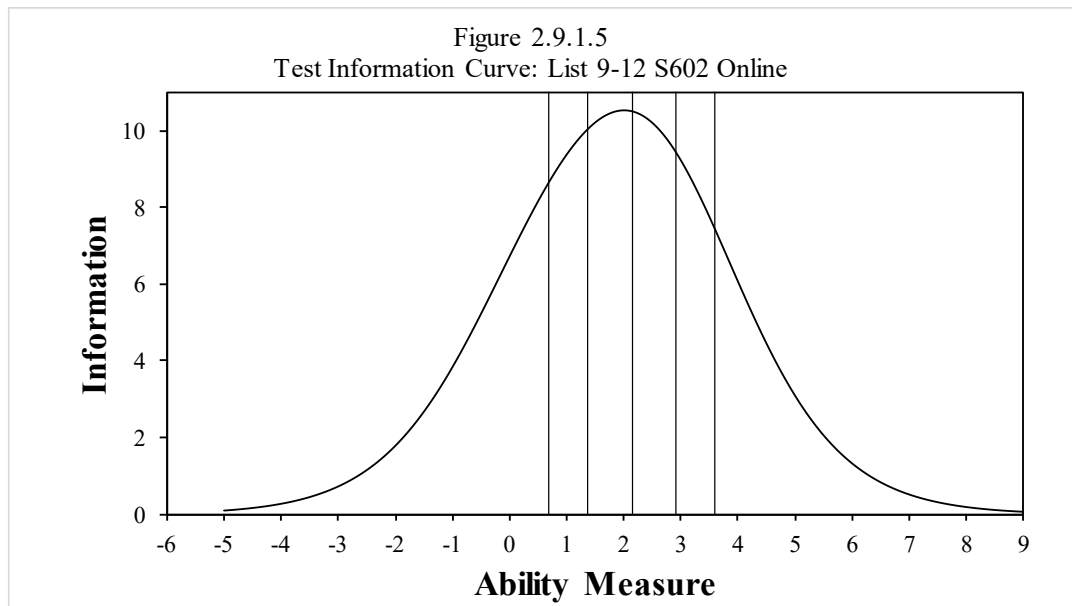
Test Information Curve: List 6–8 S602 Online



2.9.1.5 Grades 9–12

Figure 2.9.1.5

Test Information Curve: List 9–12 S602 Online

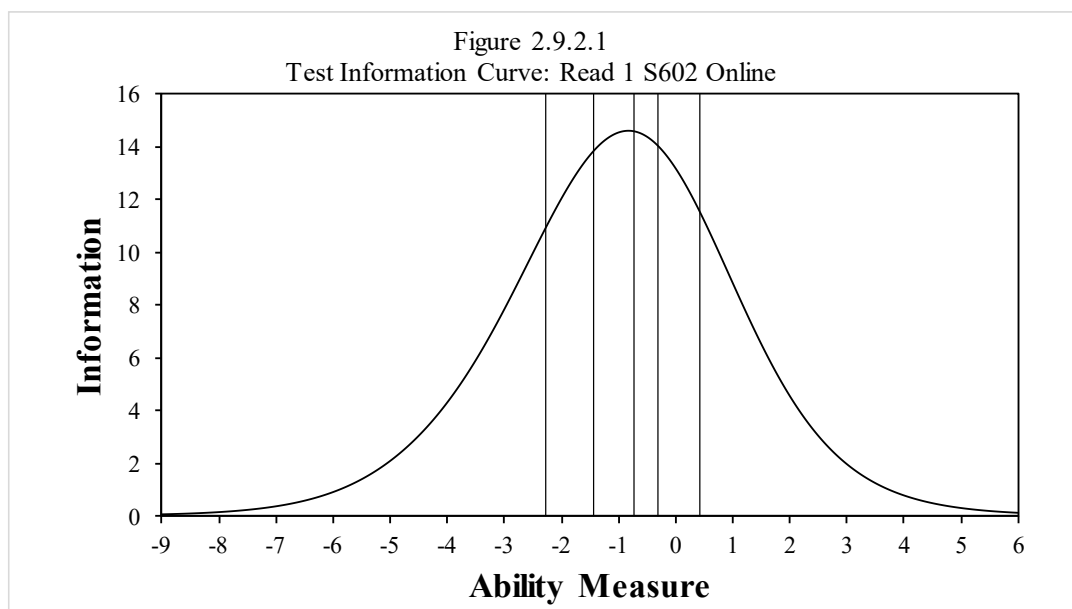


2.9.2 Reading

2.9.2.1 Grade 1

Figure 2.9.2.1

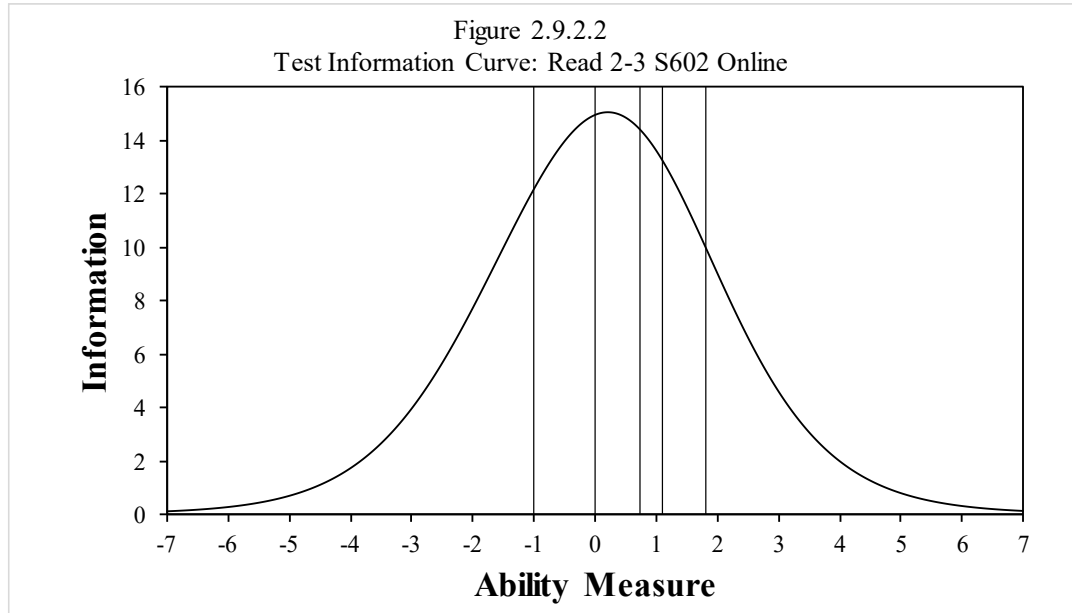
Test Information Curve: Read 1 S602 Online



2.9.2.2 Grades 2–3

Figure 2.9.2.2

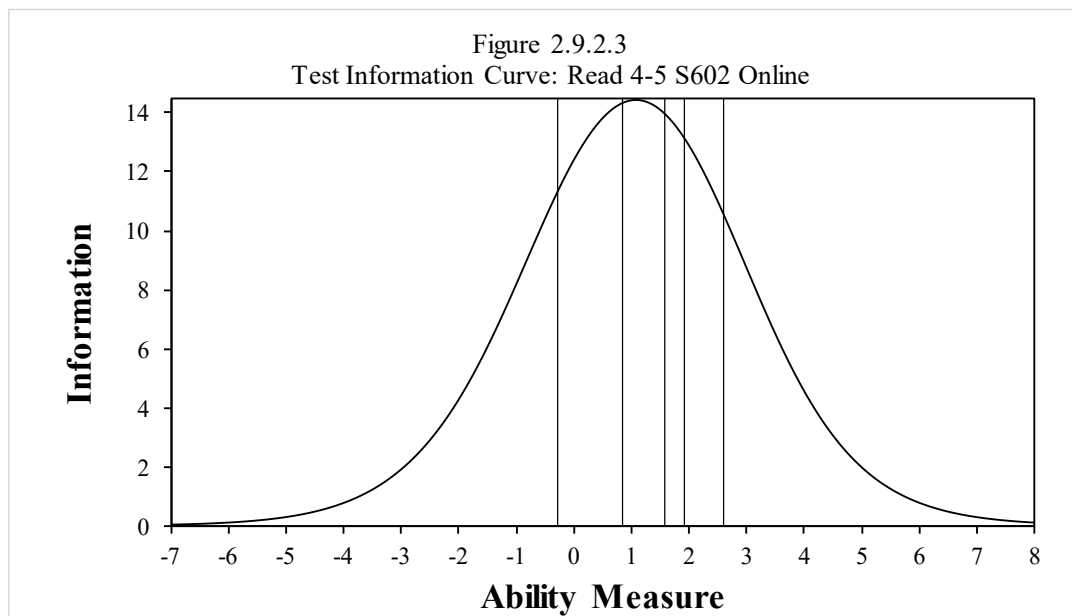
Test Information Curve: Read 2–3 S602 Online



2.9.2.3 Grades 4–5

Figure 2.9.2.3

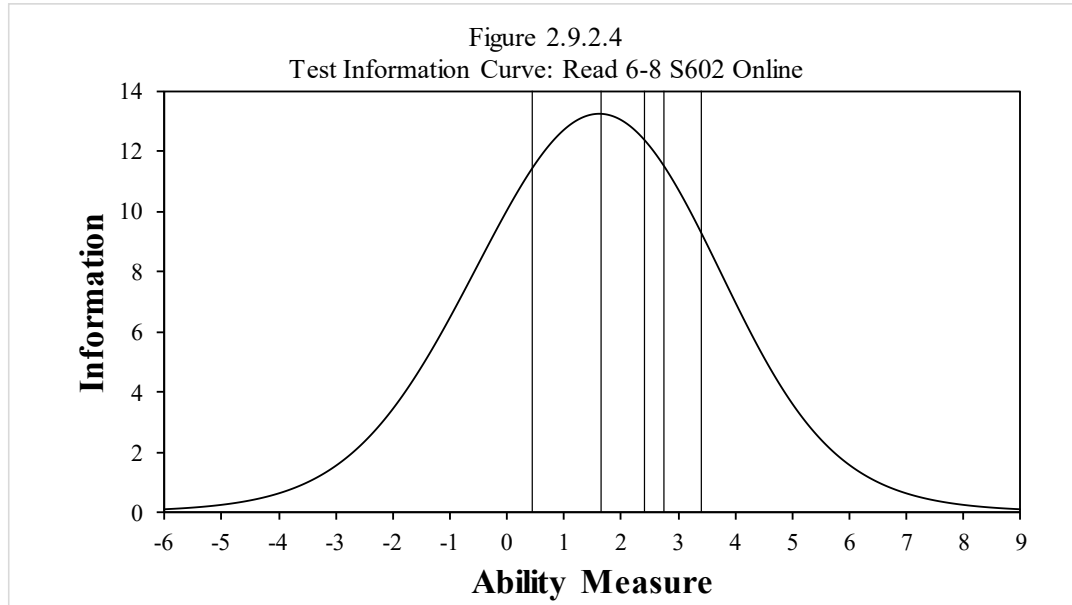
Test Information Curve: Read 4–5 S602 Online



2.9.2.4 Grades 6–8

Figure 2.9.2.4

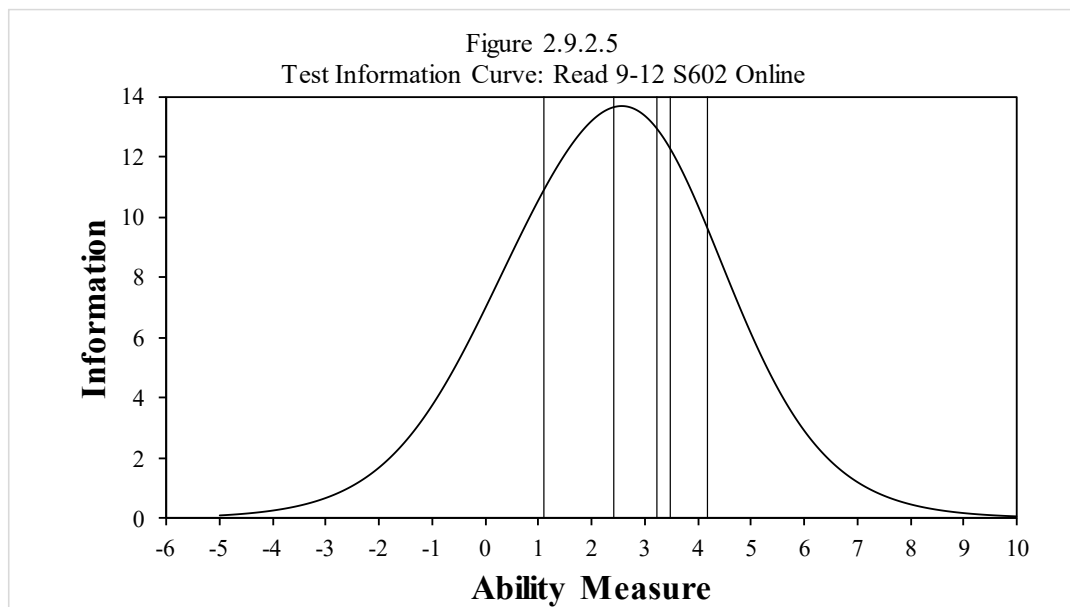
Test Information Curve: Read 6–8 S602 Online



2.9.2.5 Grades 9–12

Figure 2.9.2.5

Test Information Curve: Read 9–12 S602 Online



2.9.3 Writing

2.9.3.1 Grade 1

Figure 2.9.3.1.1

Test Information Curve: Writ 1 A S602 Online

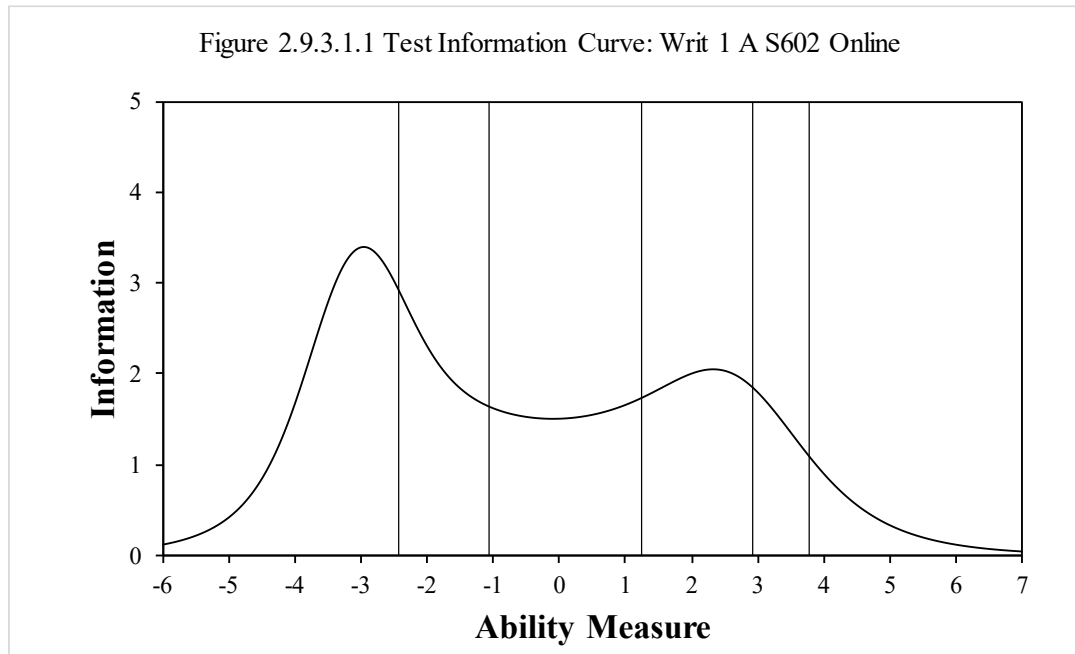


Figure 2.9.3.1.2

Test Information Curve: Writ 1 B/C S602 Online

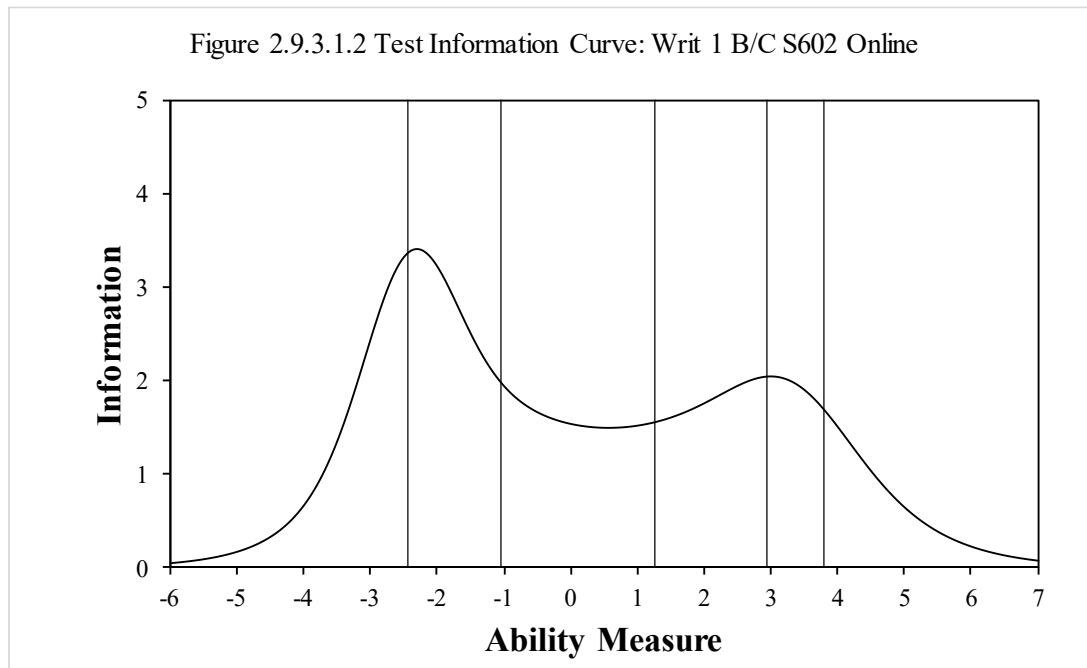
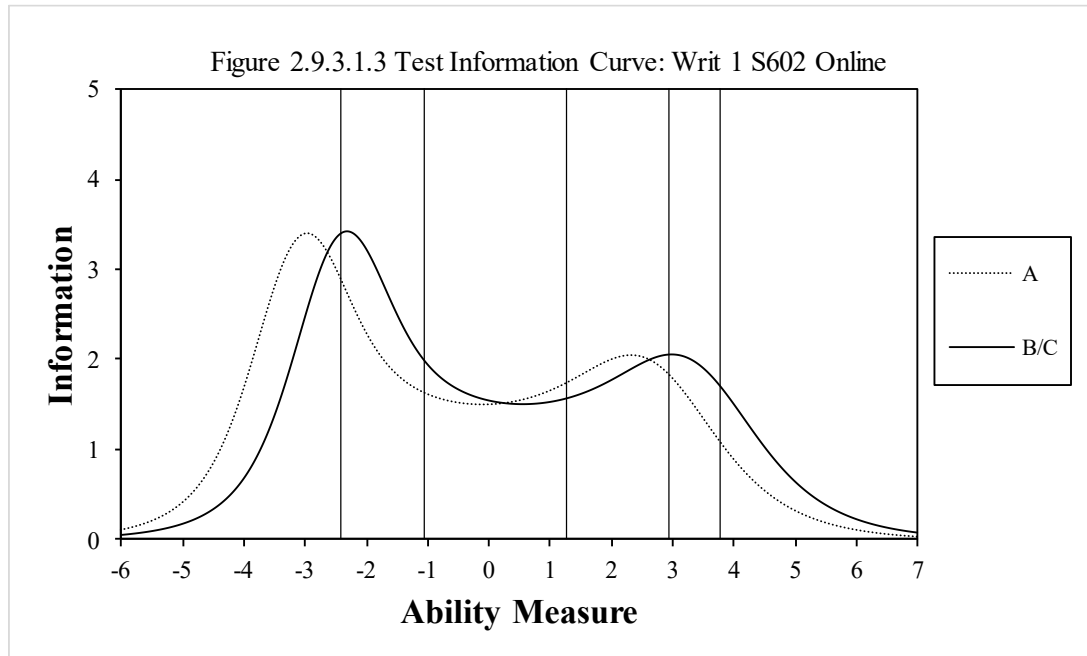


Figure 2.9.3.1.3

Test Information Curve: Writ 1 S602 Online



2.9.3.2 Grades 2–3

Figure 2.9.3.2.1

Test Information Curve: Writ 2–3 A S602 Online

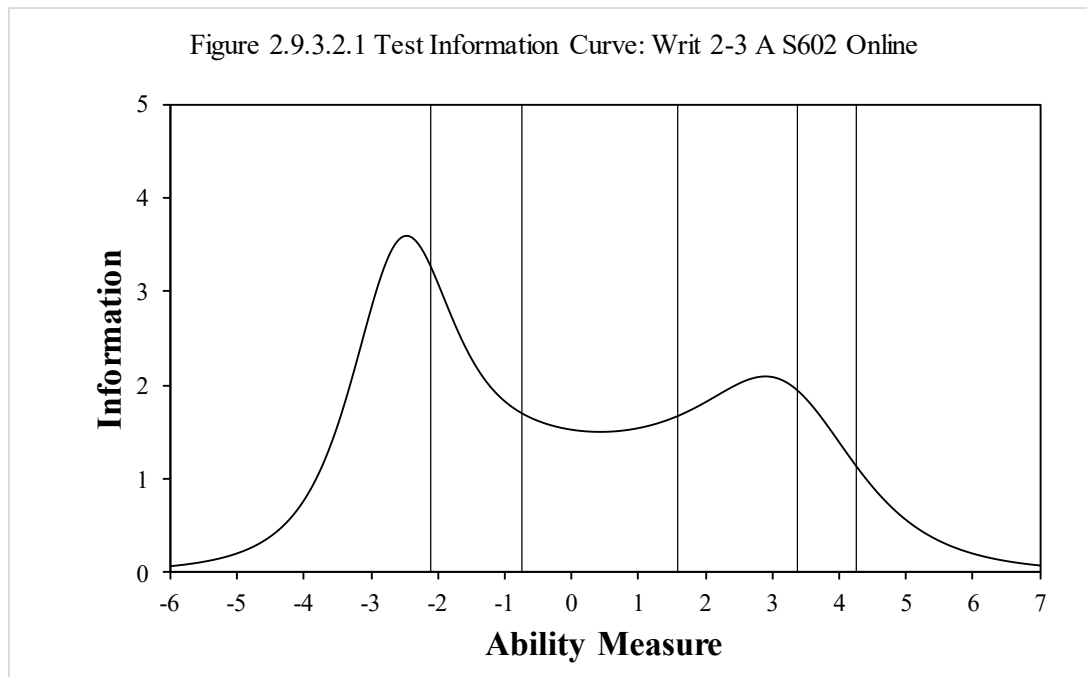


Figure 2.9.3.2.2

Test Information Curve: Writ 2–3 B/C S602 Online

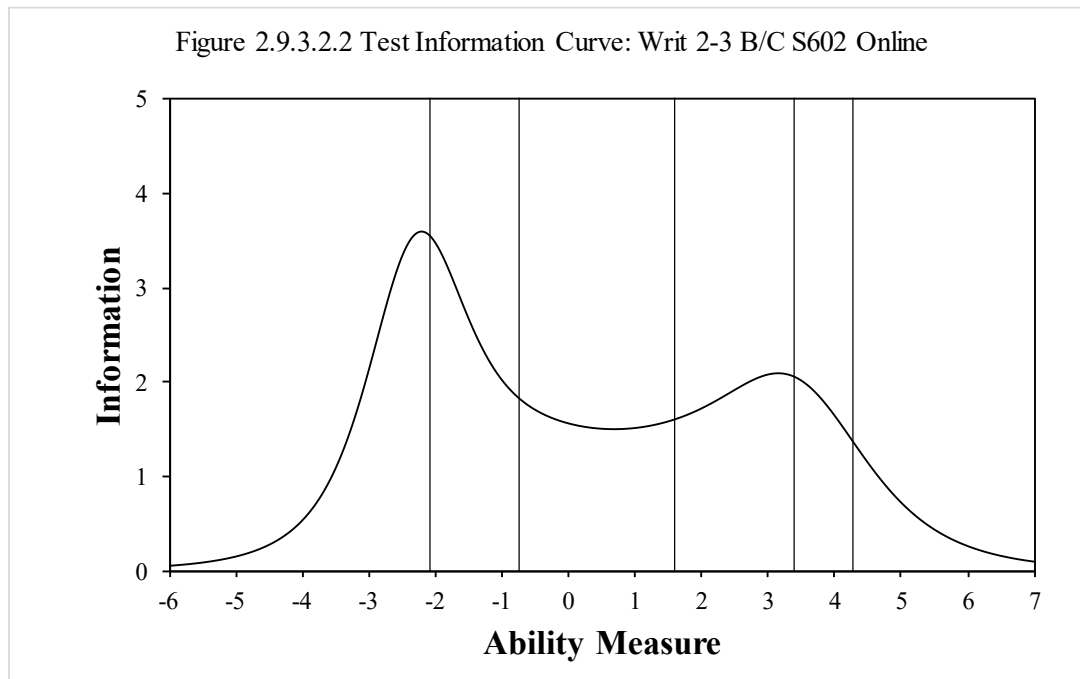
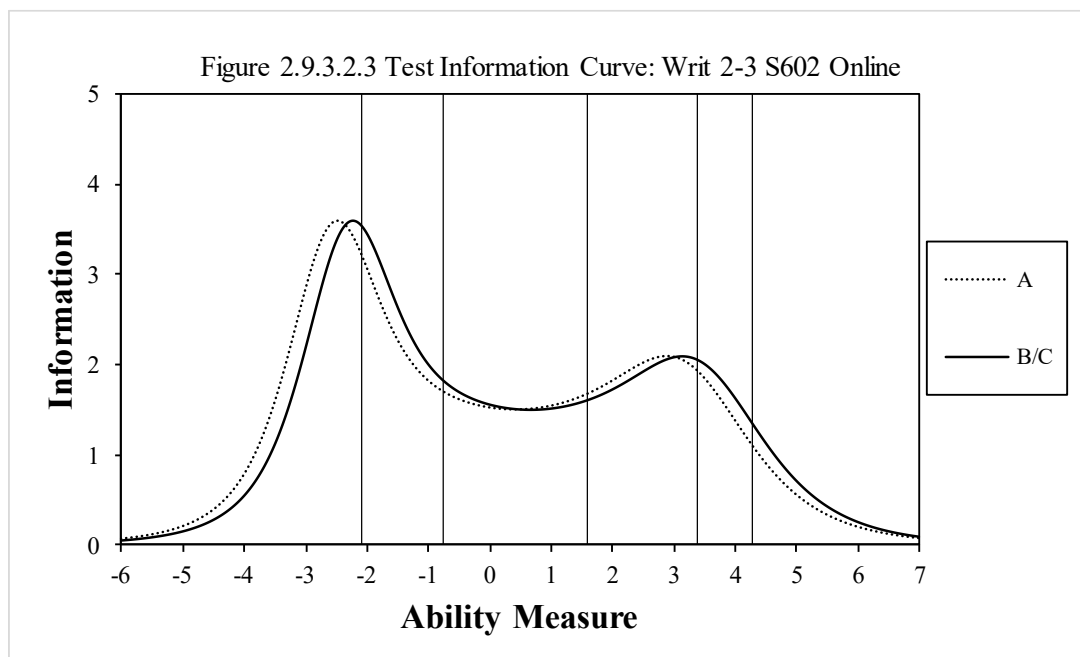


Figure 2.9.3.2.3

Test Information Curve: Writ 2–3 S602 Online



2.9.3.3 Grades 4–5

Figure 2.9.3.3.1

Test Information Curve: Writ 4–5 A S602 Online

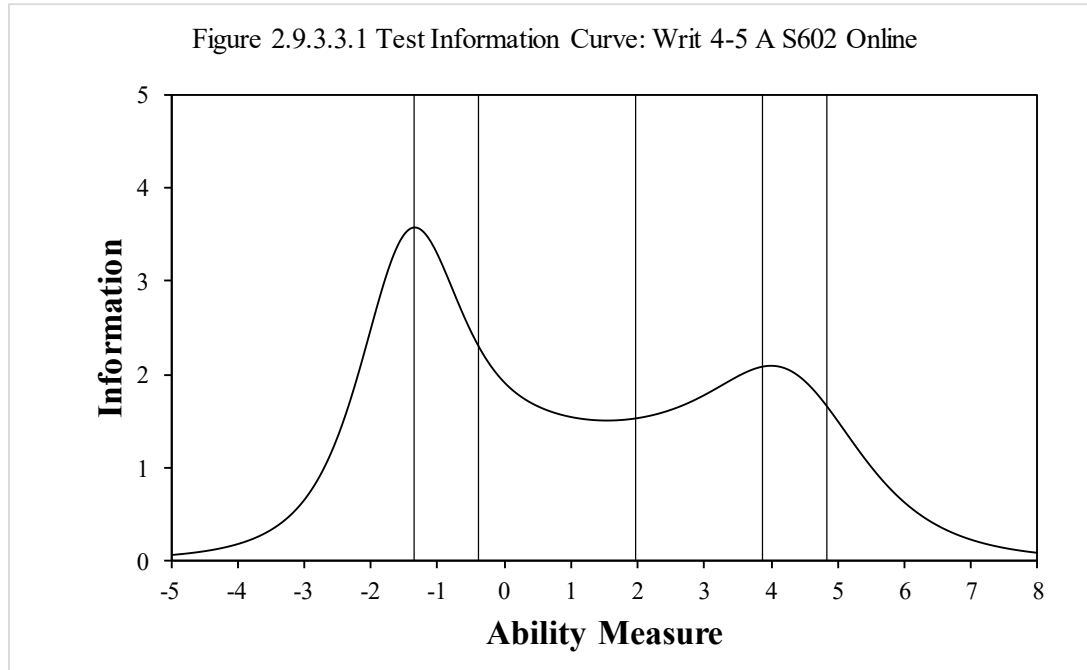


Figure 2.9.3.3.2

Test Information Curve: Writ 4–5 B/C S602 Online

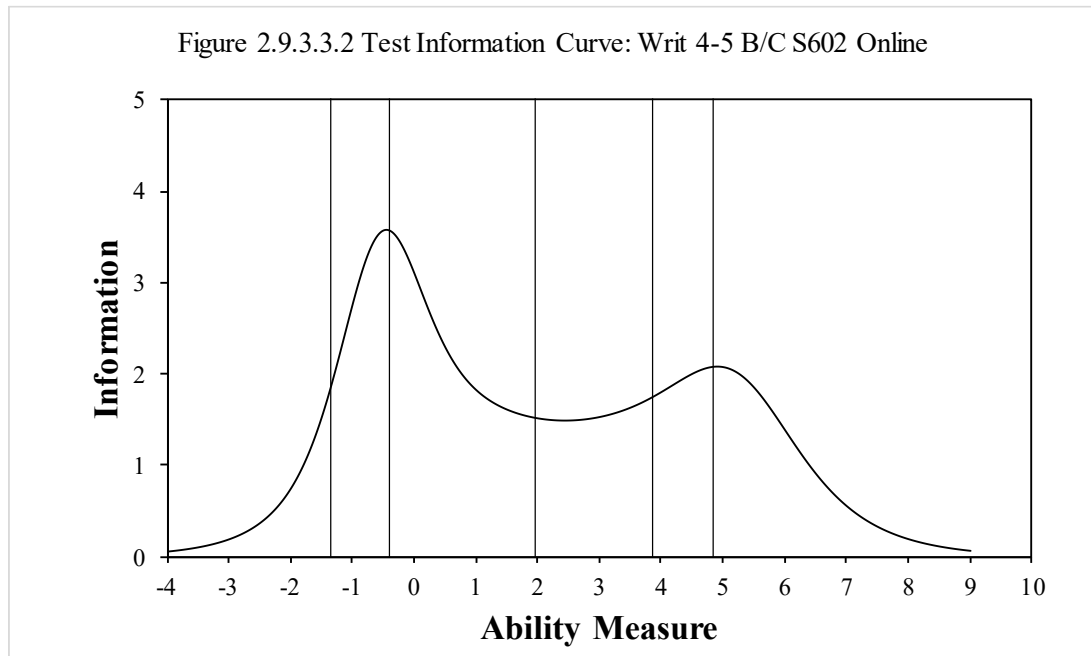
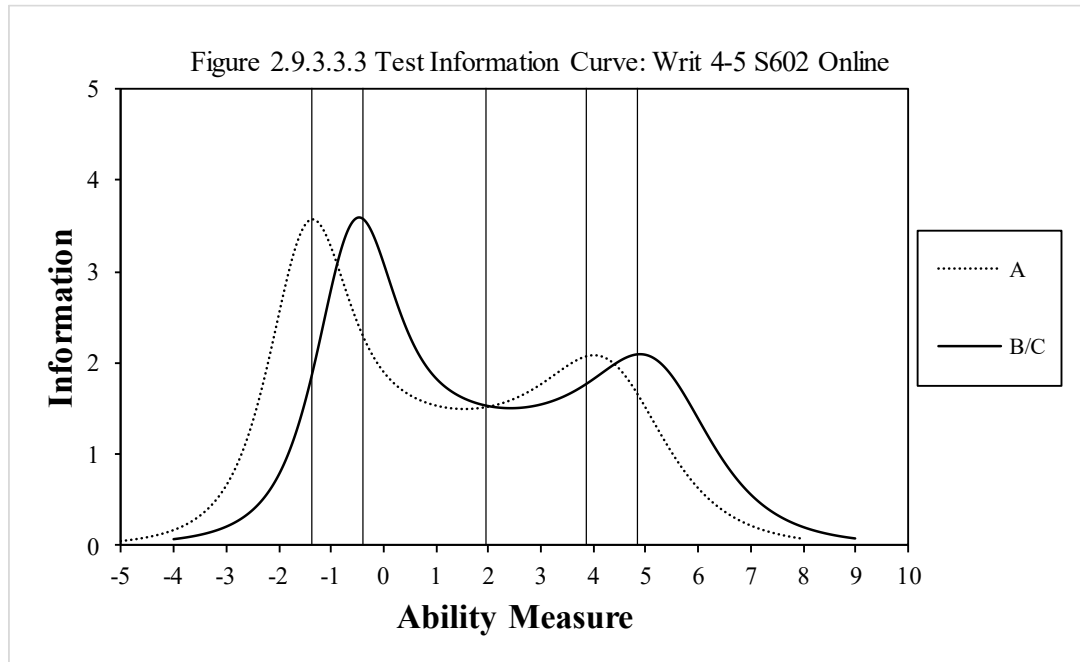


Figure 2.9.3.3.3

Test Information Curve: Writ 4–5 S602 Online



2.9.3.4 Grades 6–8

Figure 2.9.3.4.1

Test Information Curve: Writ 6–8 A S602 Online

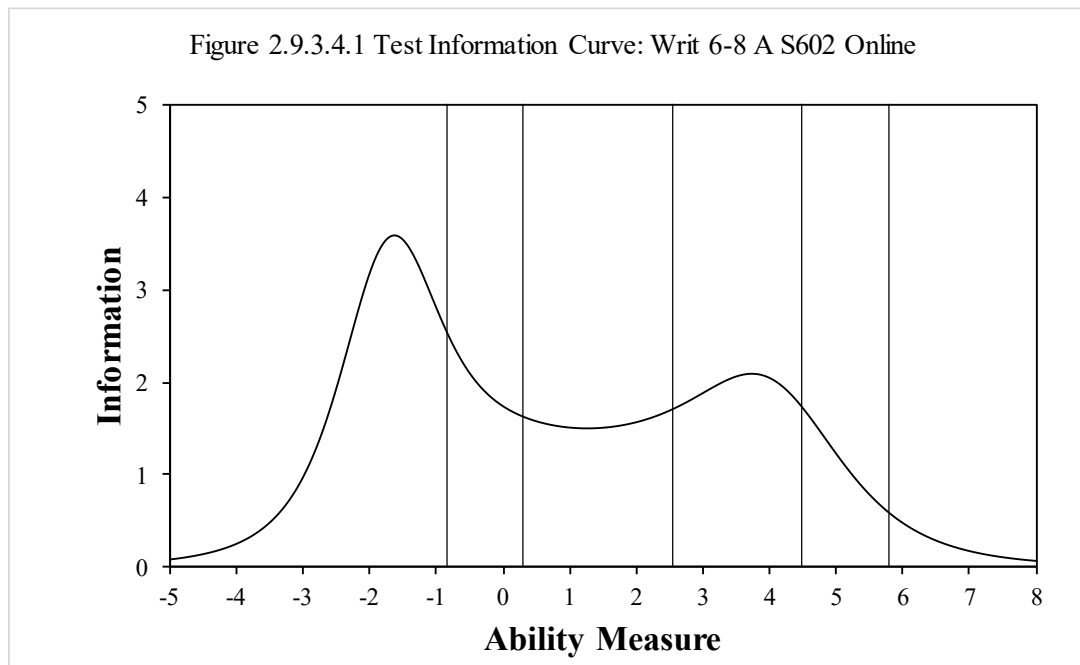


Figure 2.9.3.4.2

Test Information Curve: Writ 6–8 B/C S602 Online

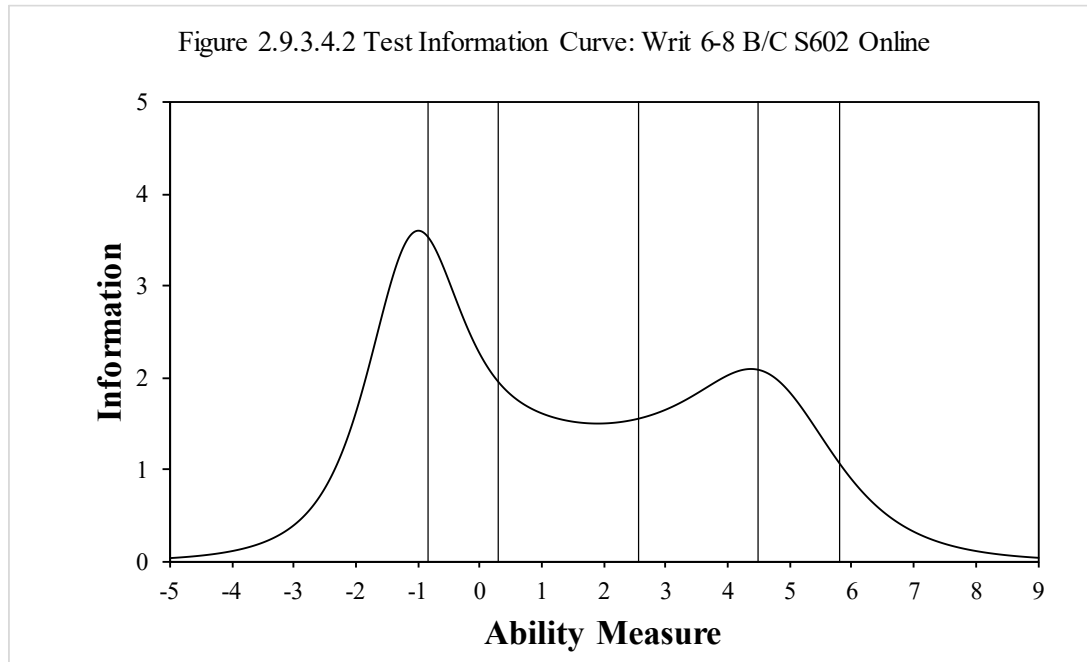
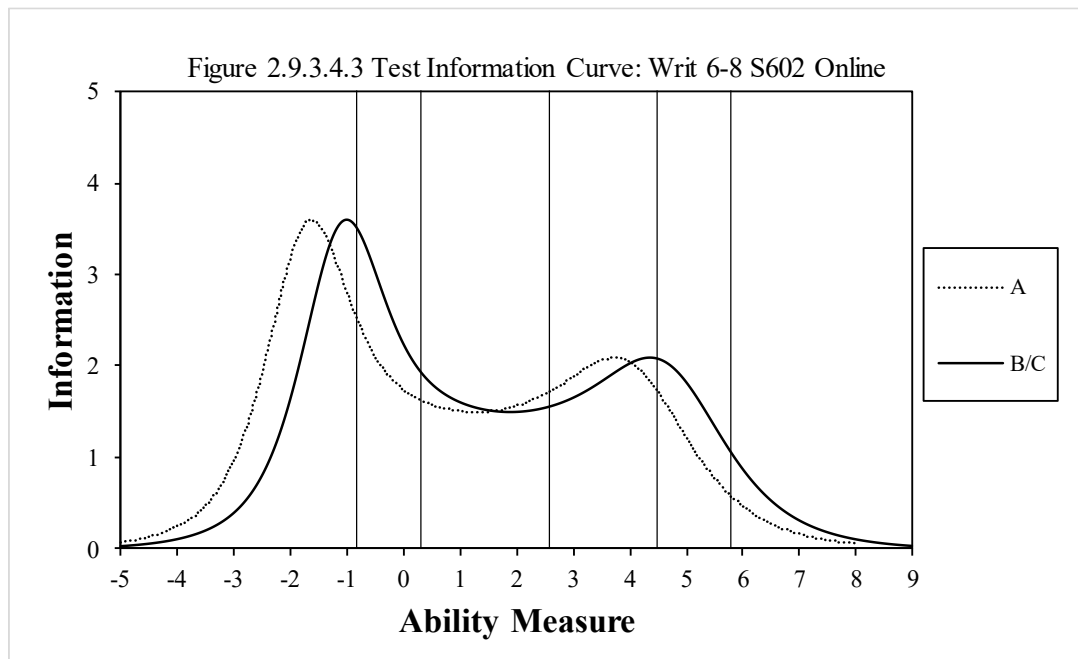


Figure 2.9.3.4.3

Test Information Curve: Writ 6–8 S602 Online



2.9.3.5 Grades 9–12

Figure 2.9.3.5.1

Test Information Curve: Writ 9–12 S602 Online

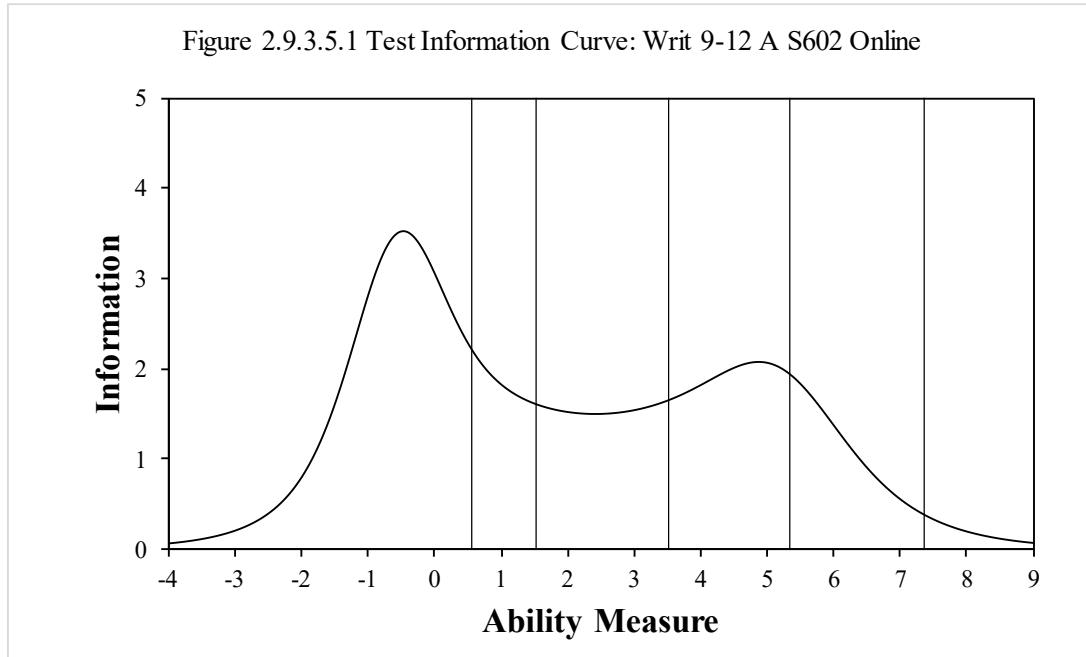


Figure 2.9.3.5.2

Test Information Curve: Writ 9–12 B/C S602 Online

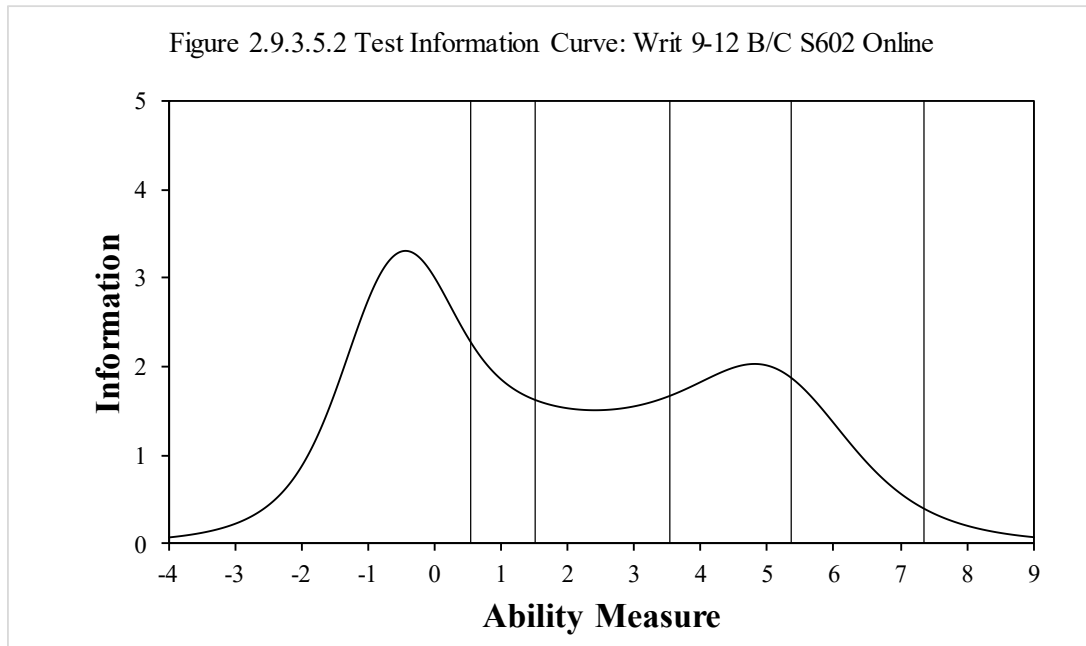
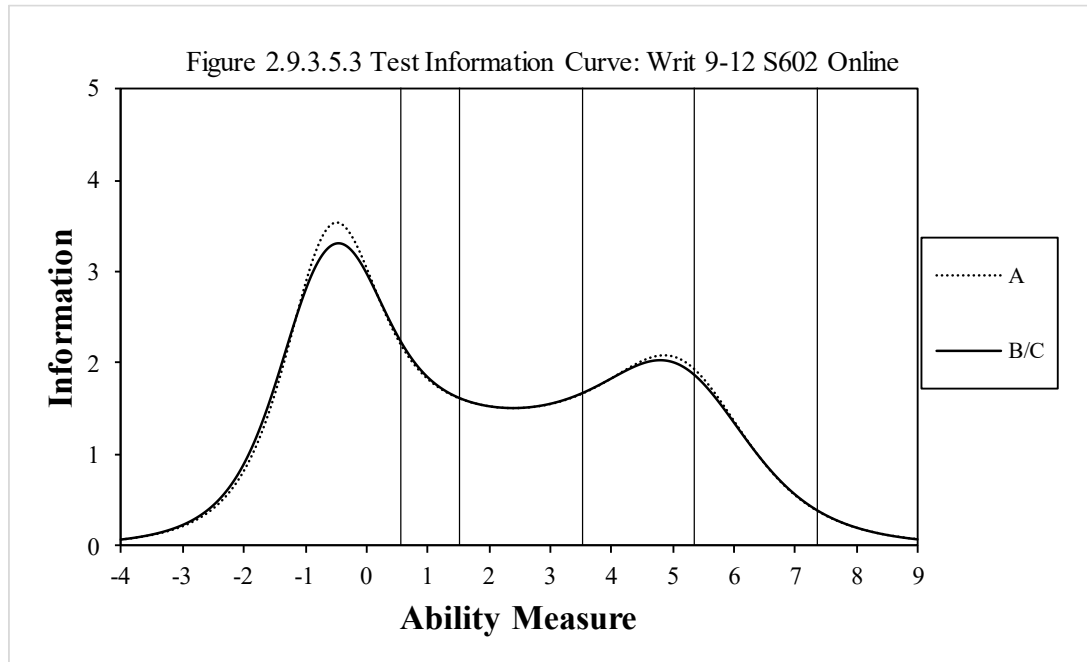


Figure 2.9.3.5.3

Test Information Curve: Writ 9–12 S 602 Online



2.9.4 Speaking

2.9.4.1 Grade 1

Figure 2.9.4.1.1

Test Information Curve: Spek 1 Pre-A S602 Online

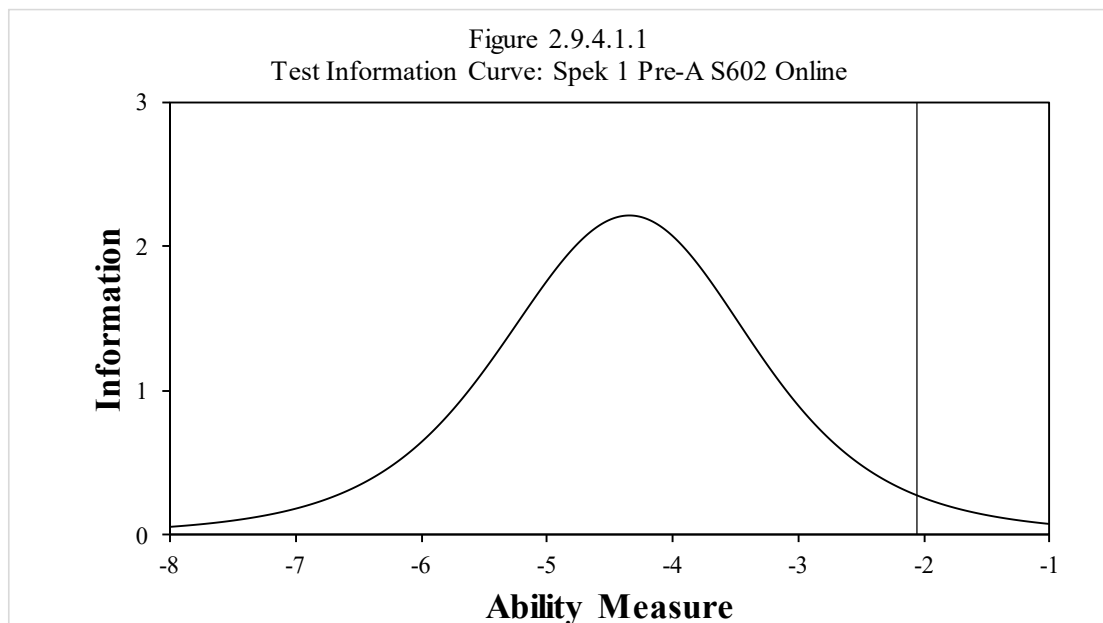


Figure 2.9.4.1.2

Test Information Curve: Spek 1 A S602 Online

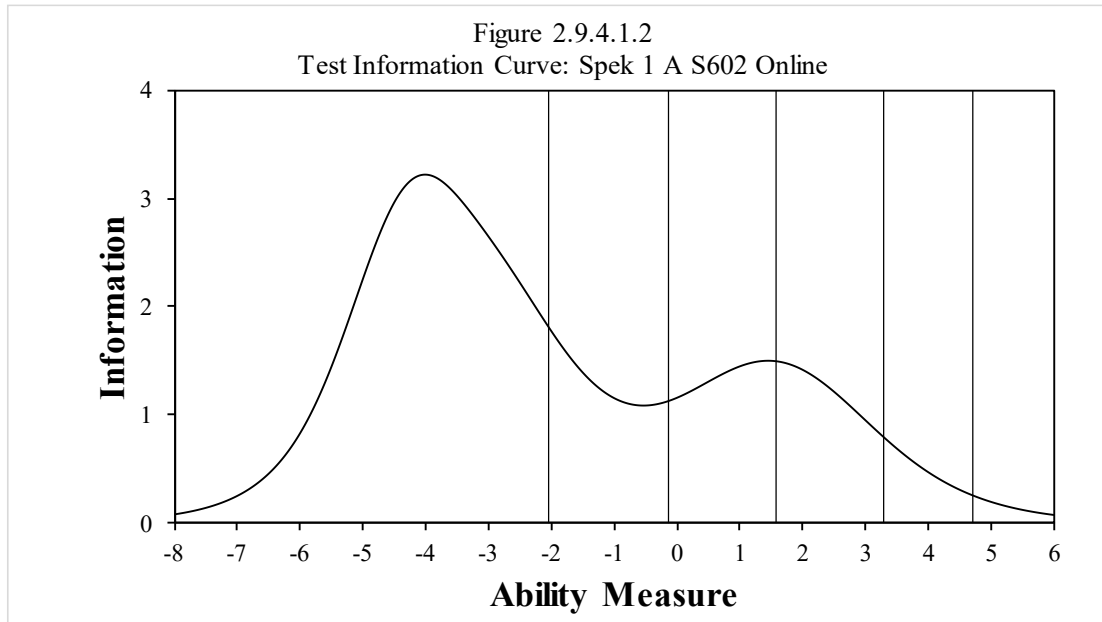


Figure 2.9.4.1.3

Test Information Curve: Spek 1 B/C S602 Online

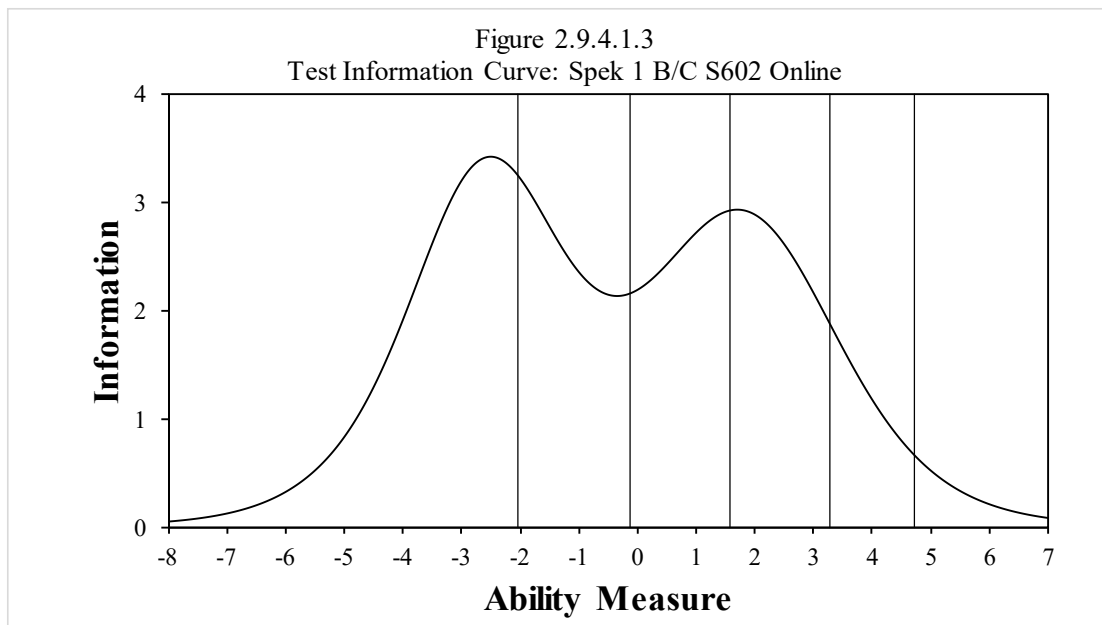
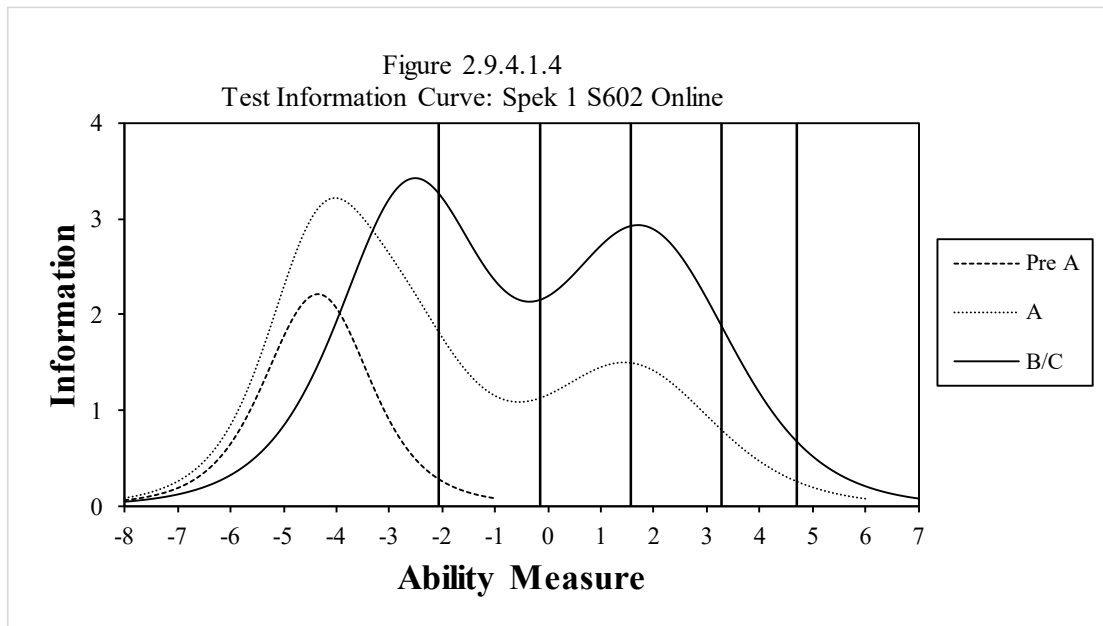


Figure 2.9.4.1.4

Test Information Curve: Spek 1 S602 Online



2.9.4.2 Grades 2–3

Figure 2.9.4.2.1

Test Information Curve: Spek 2–3 Pre-A S602 Online

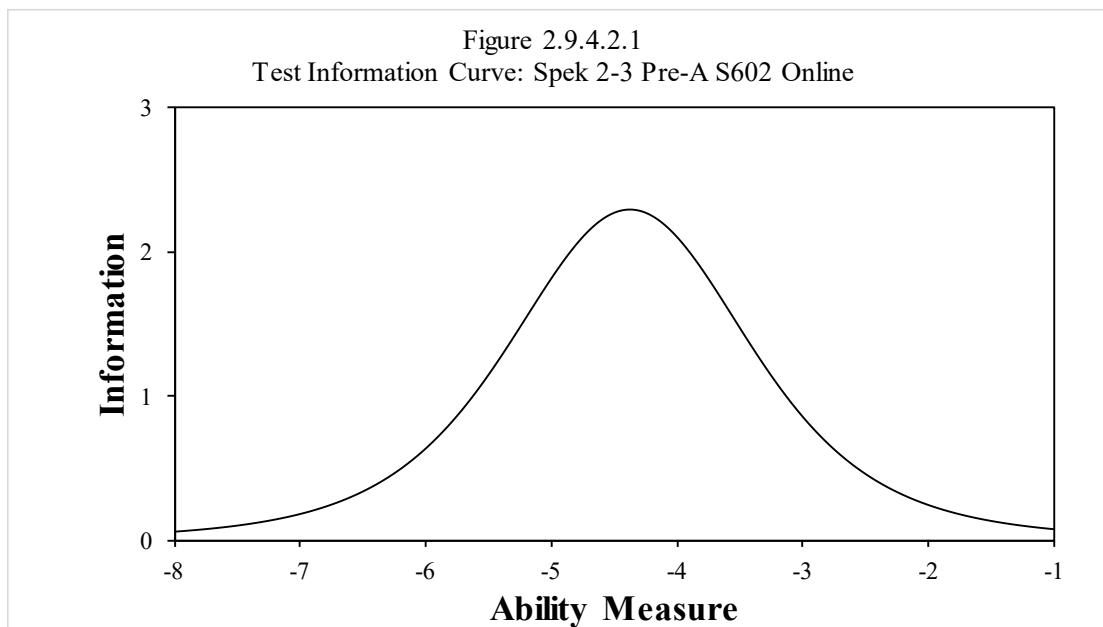


Figure 2.9.4.2.2

Test Information Curve: Spek 2-3 A S602 Online

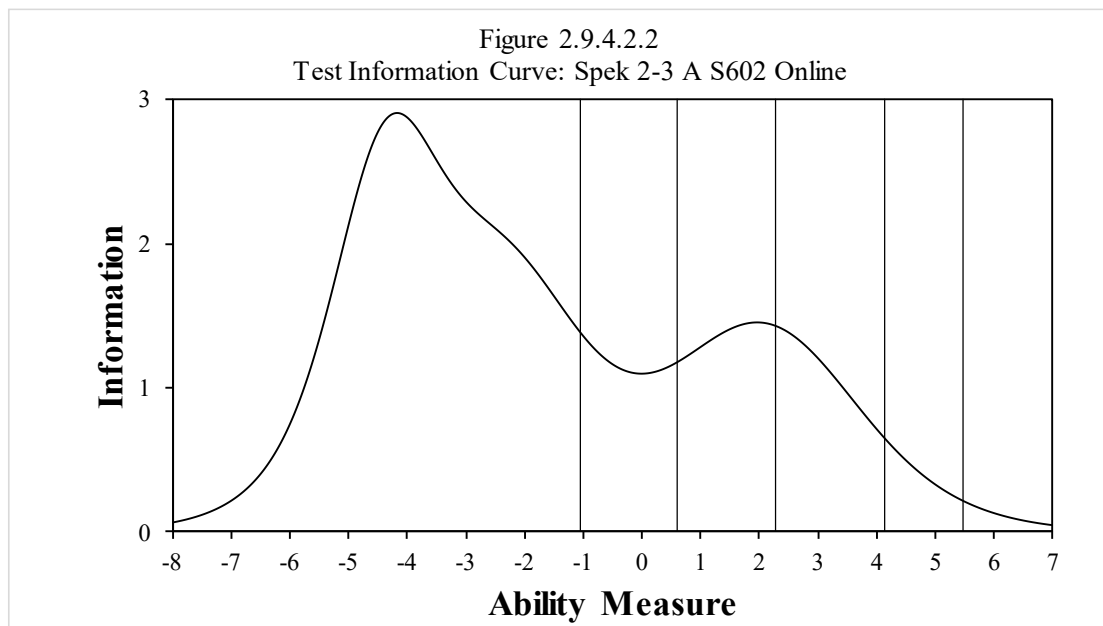


Figure 2.9.4.2.3

Test Information Curve: Spek 2-3 B/C S602 Online

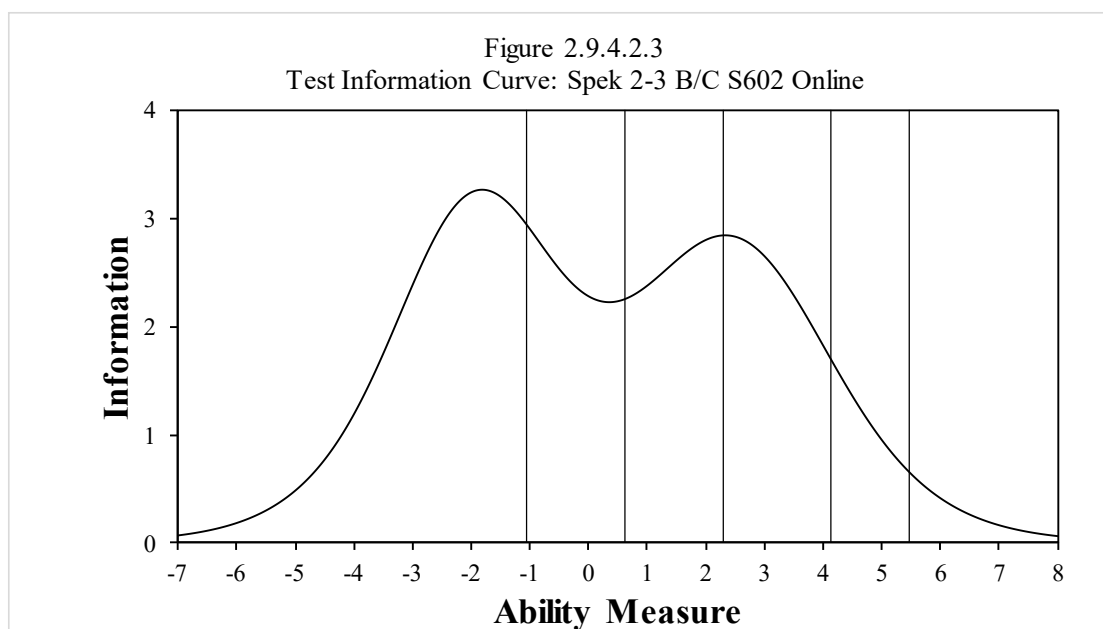
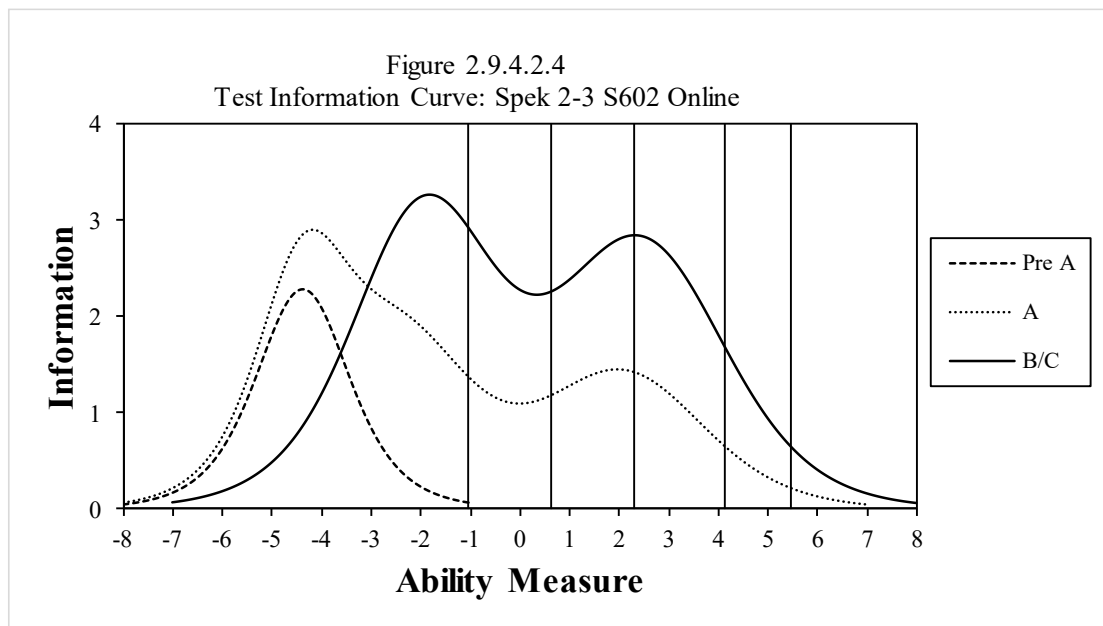


Figure 2.9.4.2.4

Test Information Curve: Spek 2-3 S602 Online



2.9.4.3 Grades 4-5

Figure 2.9.4.3.1

Test Information Curve: Spek 4-5 Pre-A S602 Online

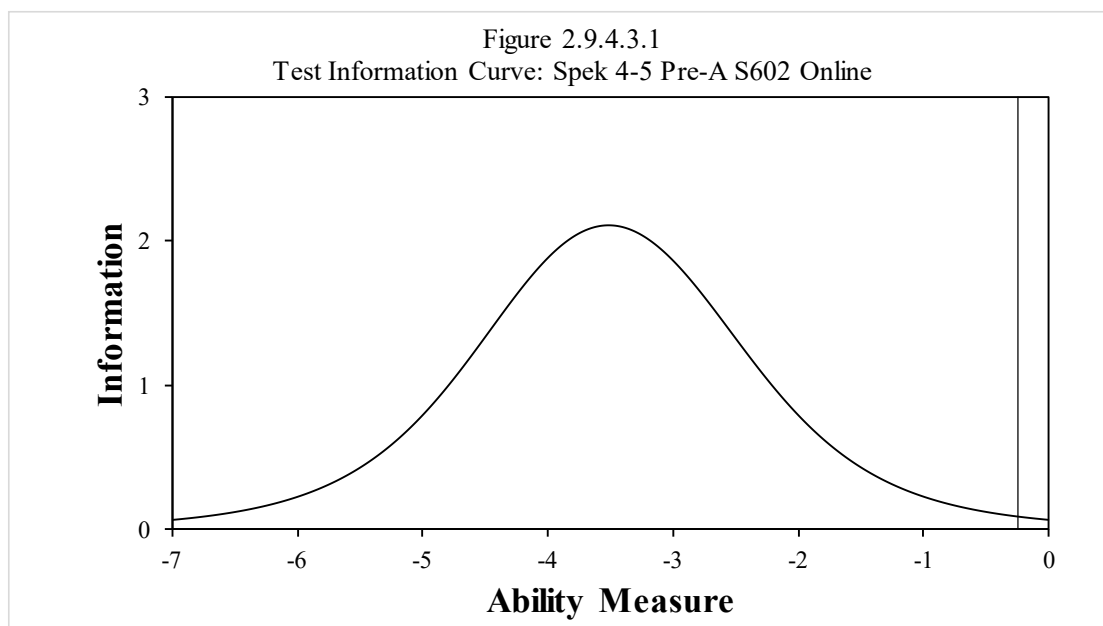


Figure 2.9.4.3.2

Test Information Curve: Spek 4-5 A S602 Online

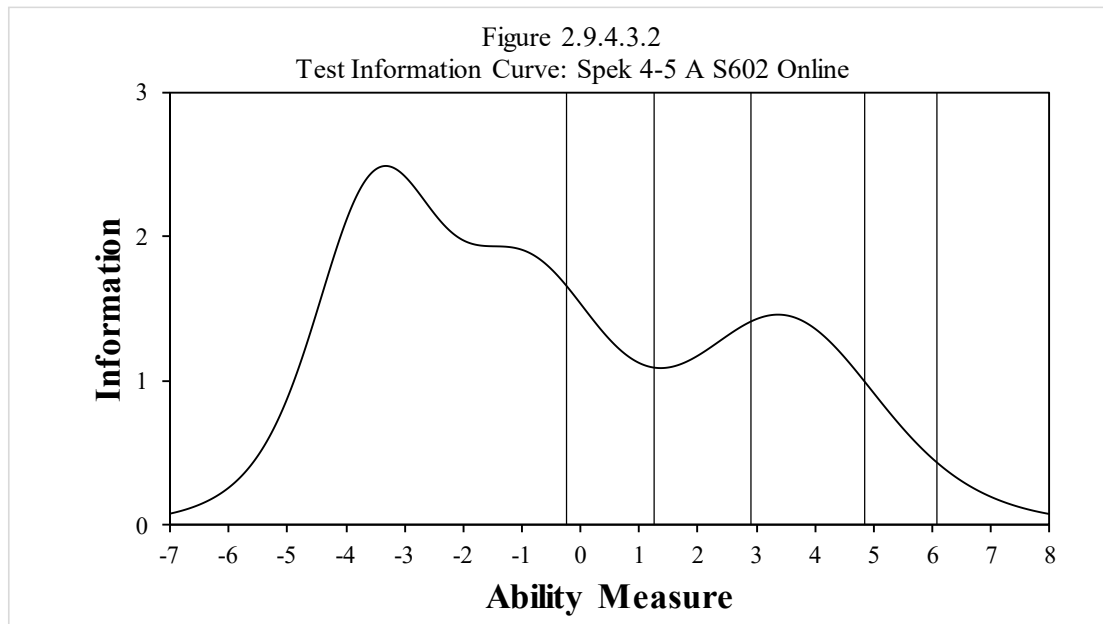


Figure 2.9.4.3.3

Test Information Curve: Spek 4-5 B/C S602 Online

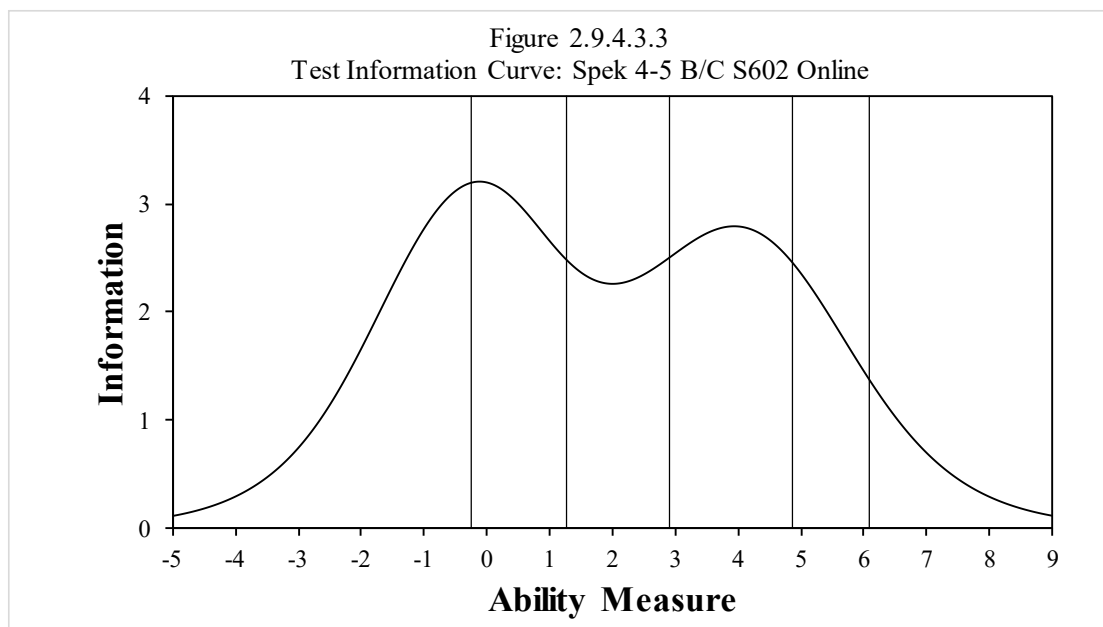
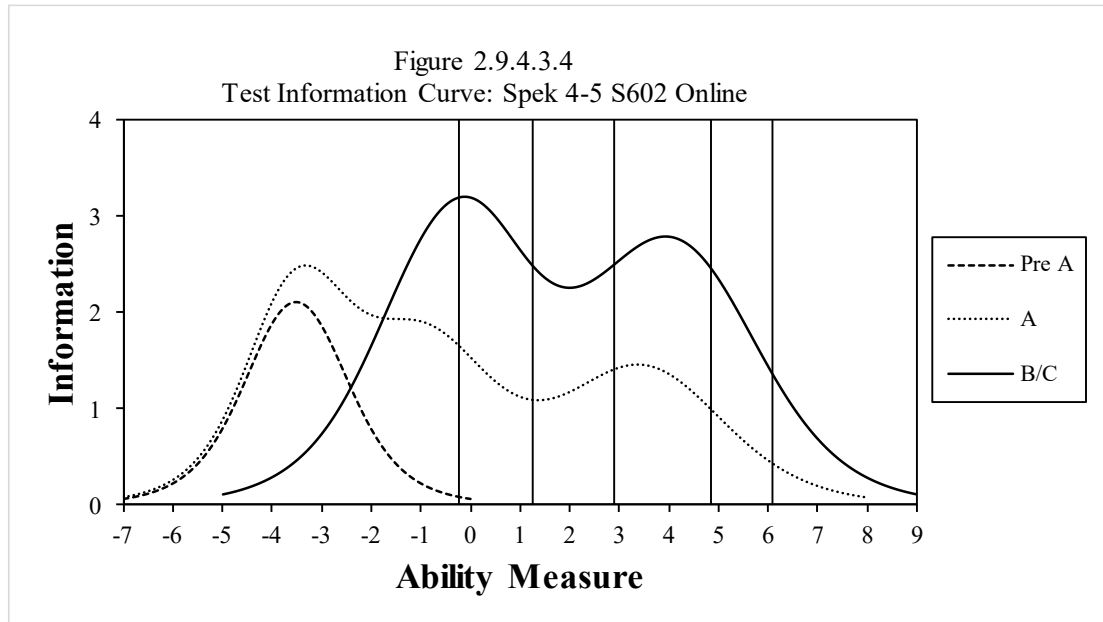


Figure 2.9.4.3.4

Test Information Curve: Spek 4-5 S602 Online



2.9.4.4 Grades 6-8

Figure 2.9.4.4.1

Test Information Curve: Spek 6-8 Pre-A S602 Online

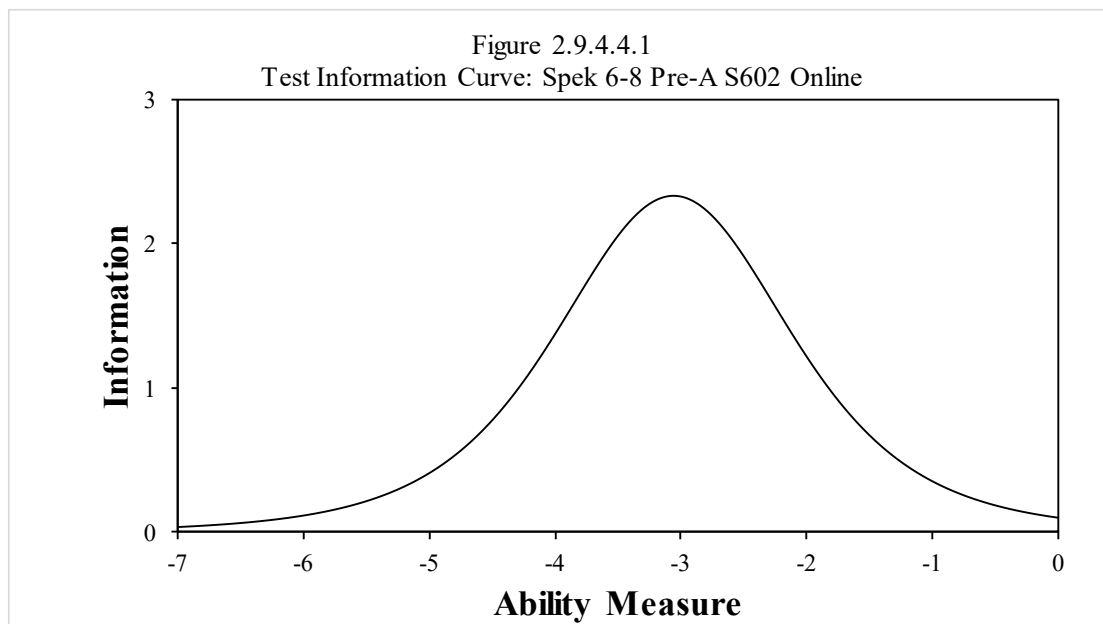


Figure 2.9.4.4.2

Test Information Curve: Spek 6–8 A S602 Online

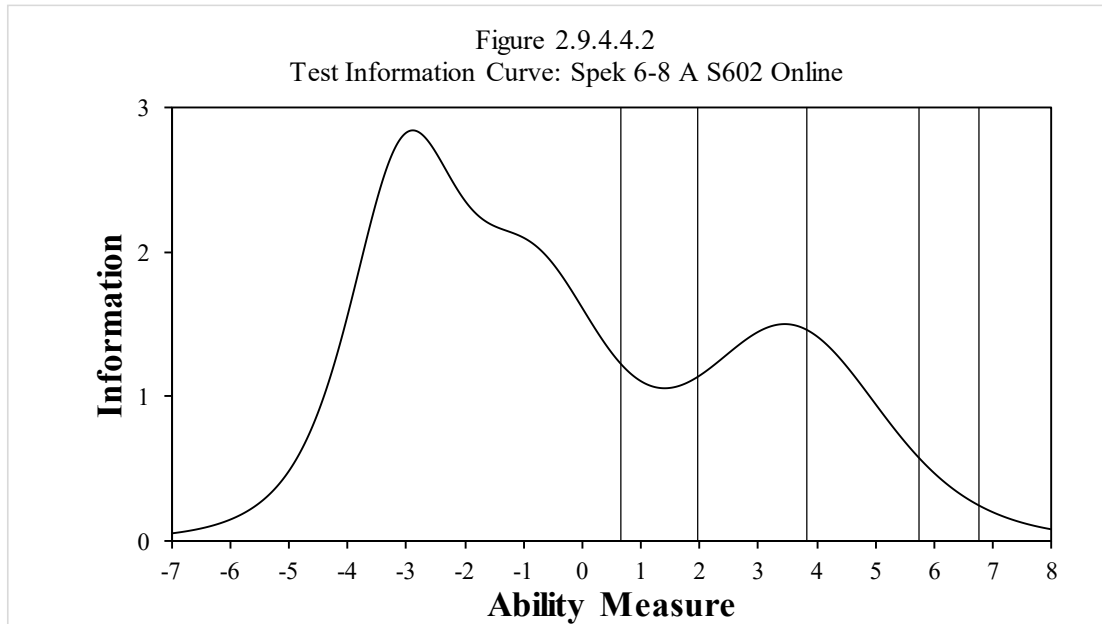


Figure 2.9.4.4.3

Test Information Curve: Spek 6–8 B/C S602 Online

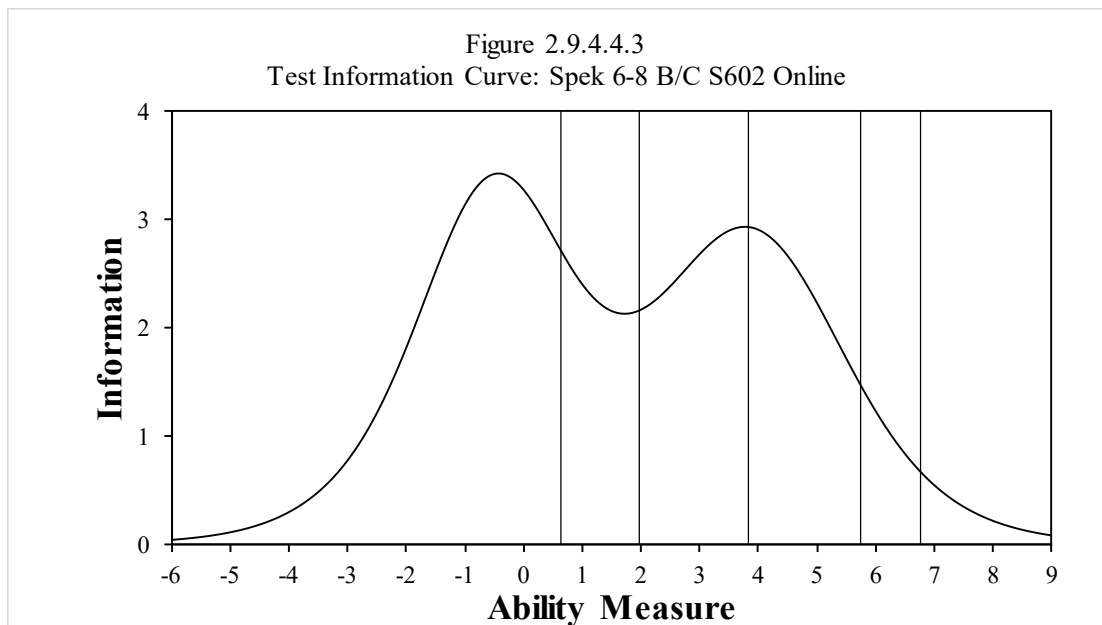
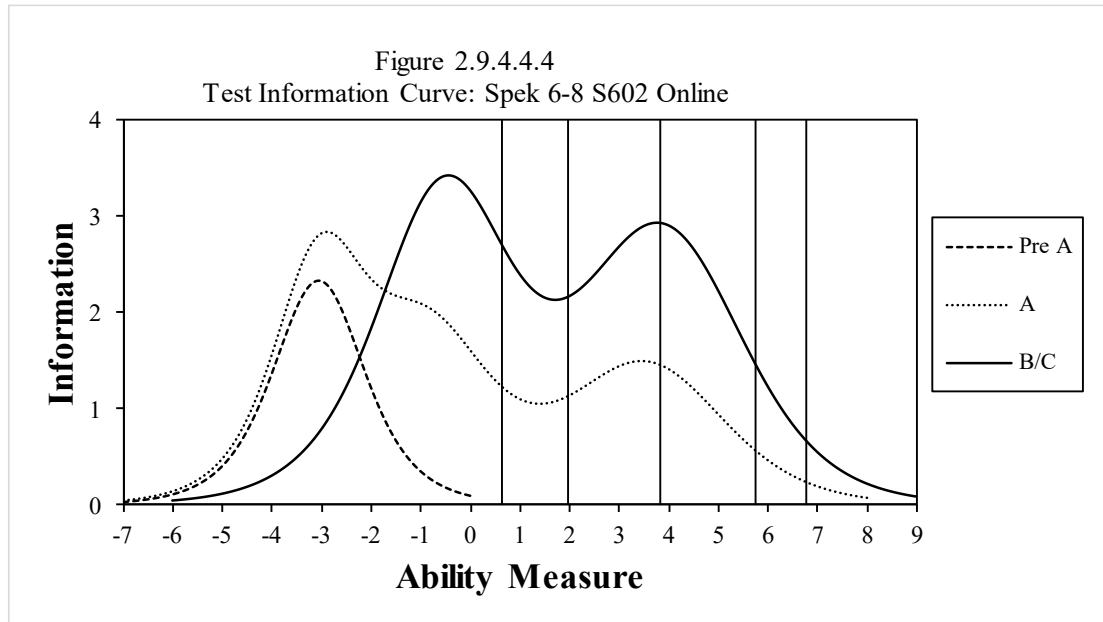


Figure 2.9.4.4.4

Test Information Curve: Spek 6–8 S602 Online



2.9.4.5 Grades 9–12

Figure 2.9.4.5.1

Test Information Curve: Spek 9–12 Pre-A S602 Online

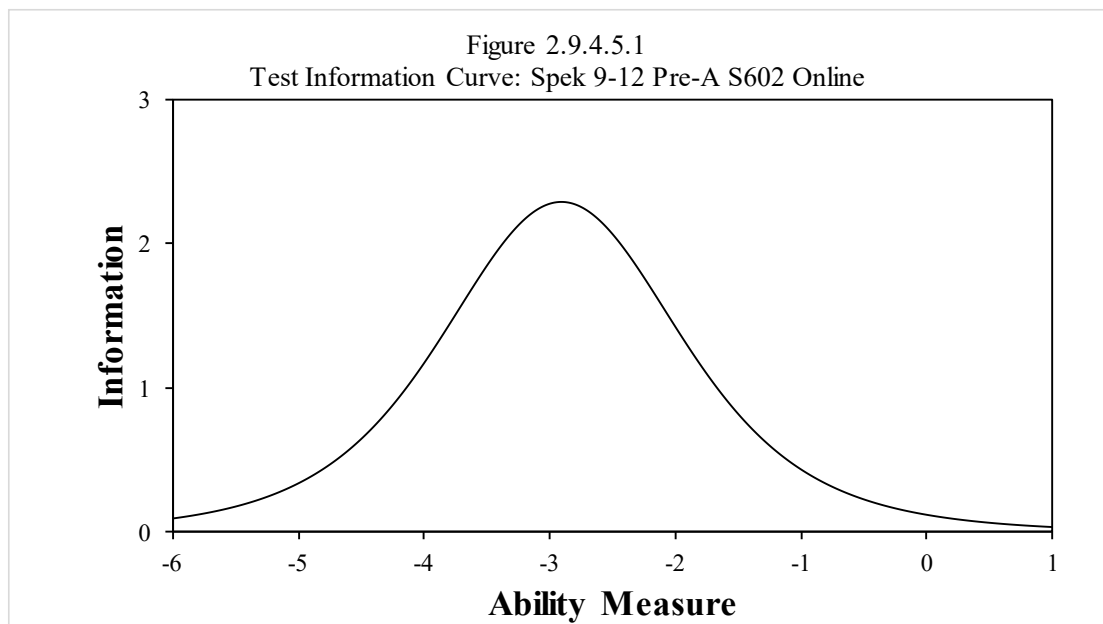


Figure 2.9.4.5.2

Test Information Curve: Spek 9–12 A S602 Online

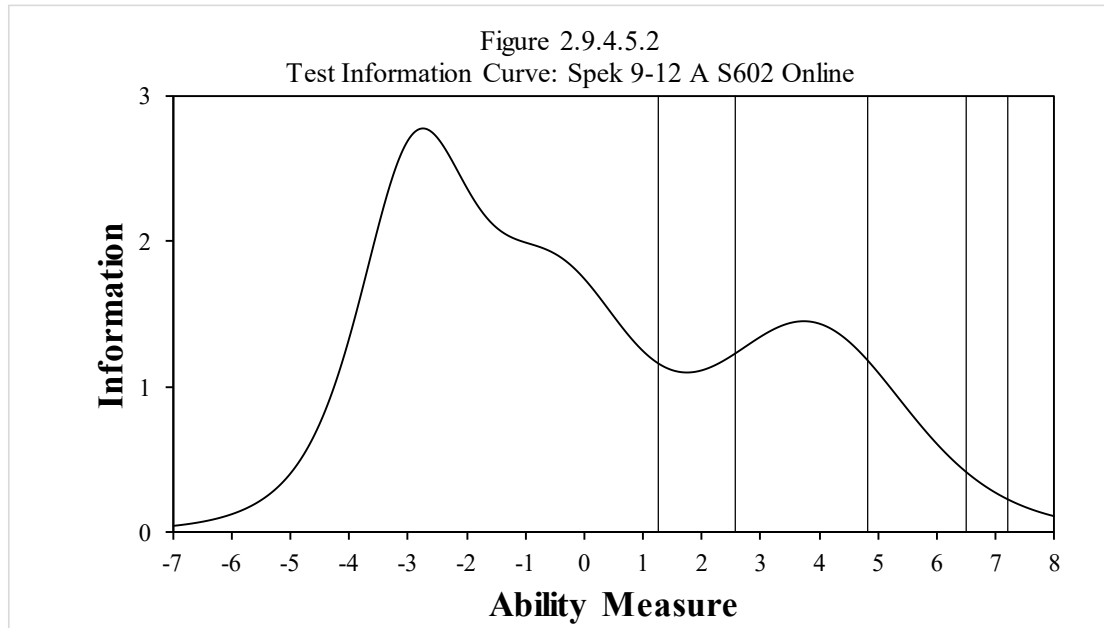


Figure 2.9.4.5.3

Test Information Curve: Spek 9–12 B/C S602 Online

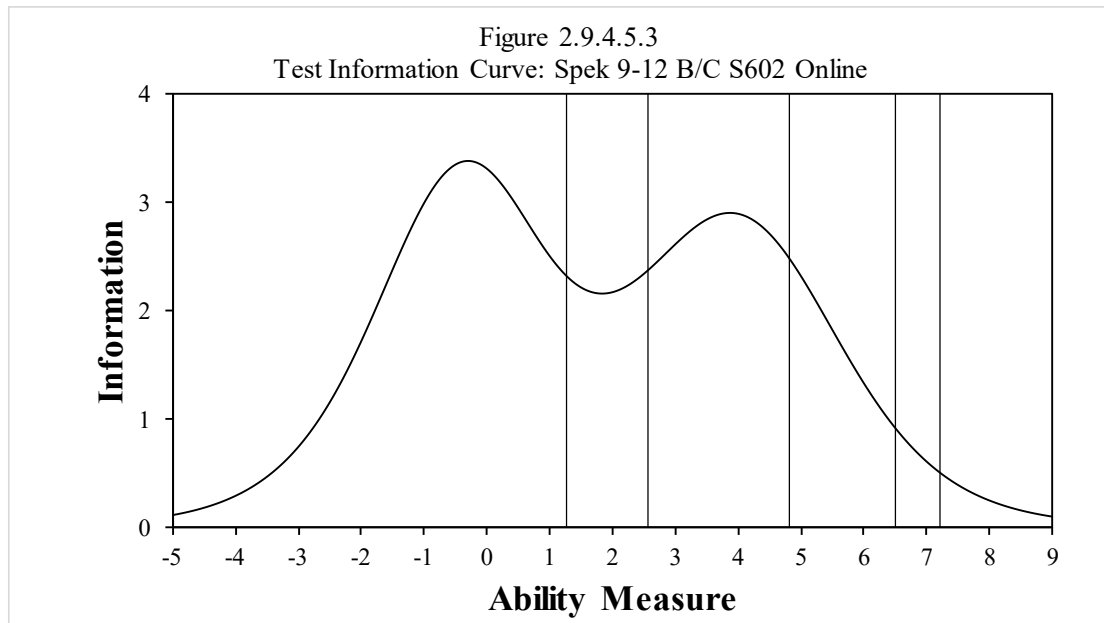
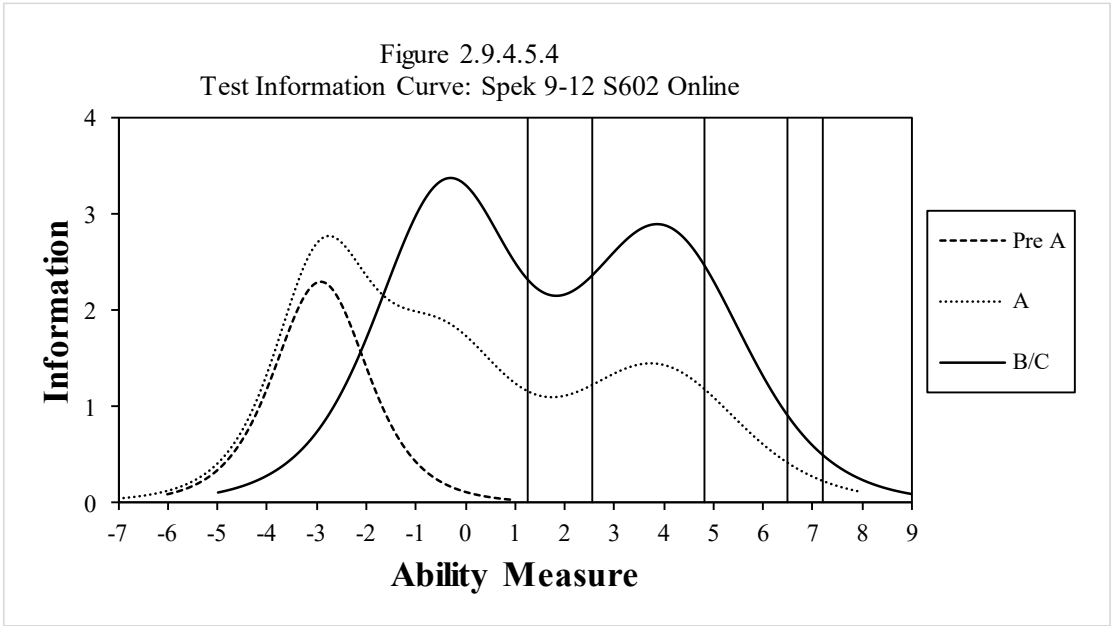


Figure 2.9.4.5.4

Test Information Curve: Spek 9-12 S602 Online



3. Analysis of Composite Scores

We calculate four composite scores for ACCESS Online: Oral Language, Literacy, Comprehension, and Overall. We calculate these composite scores as weighted averages of domain scale scores, as follows:

- Oral Language: 50% Listening + 50% Speaking
- Literacy: 50% Reading + 50% Writing
- Comprehension: 30% Listening + 70% Reading
- Overall Composite: 15% Listening + 15% Speaking + 35% Reading + 35% Writing

A policy decision by the WIDA Board, made before the first operational administration of ACCESS, resulted in the weighting, and is based on the view that literacy skills are paramount in developing academic language proficiency.

3.1 *Scale Score Distribution for Composites*

Figures and tables in this section provide scale score distributions for each of the composites, for each grade-level cluster.

For each cluster, the figure shows the distribution of the scale scores for the composite. We plotted the scale scores, grouped into units of five scale score points (e.g., 100–104, 105–109, 110–114, etc.), on the horizontal axis, and the number of students with scale scores falling into each range on the vertical axis.

Each table shows, by grade and by total for the grade-level cluster:

- The number of students in the analyses (count)
- The minimum observed scale score
- The maximum observed scale score
- The mean (average) scale score
- The standard deviation (std. dev.) of the scale score

3.1.1 Oral

3.1.1.1 Grade 1

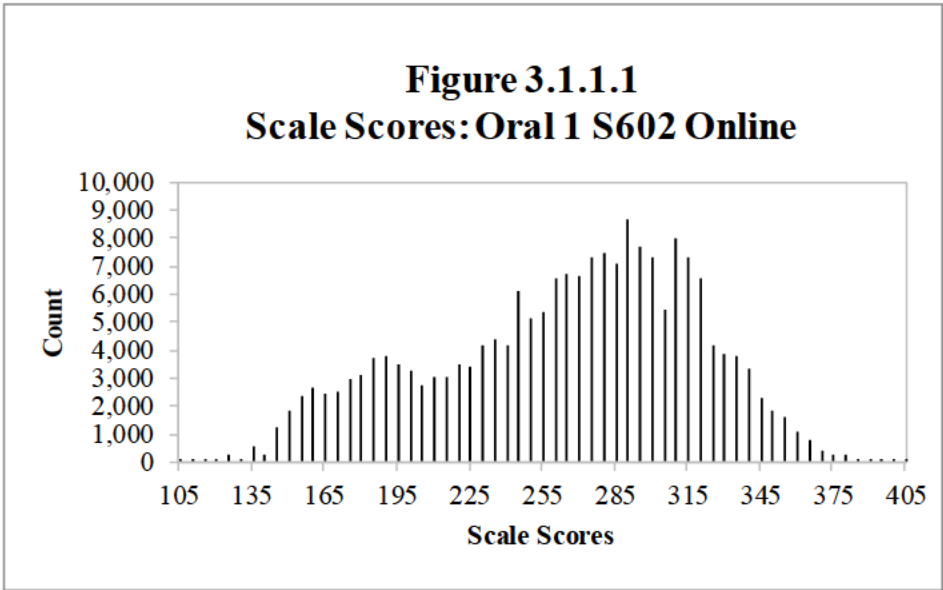
Table 3.1.1.1

Scale Score Descriptive Statistics: Oral 1 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	196,145	105	407	265.26	54.18
Total	196,145	105	407	265.26	54.18

Figure 3.1.1.1.

Scale Scores: Oral 1 S602 Online



3.1.1.2 Grades 2–3

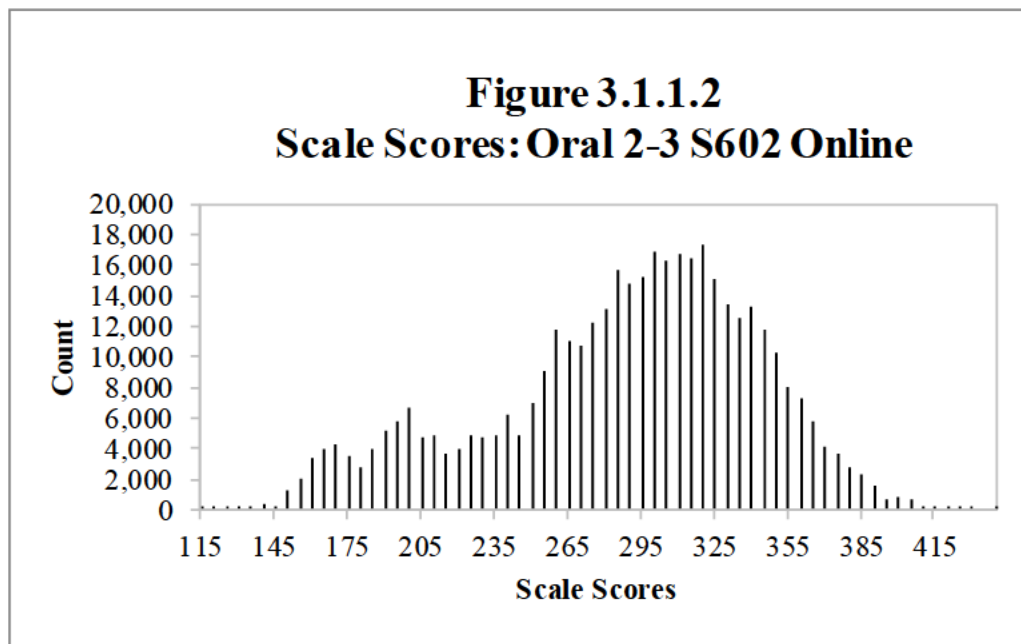
Table 3.1.1.2

Scale Score Descriptive Statistics: Oral 2–3 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	204,084	115	441	280.88	51.29
3	201,515	117	441	298.38	56.92
Total	405,599	115	441	289.57	54.86

Figure 3.1.1.2

Scale Scores: Oral 2–3 S602 Online



3.1.1.3 Grades 4–5

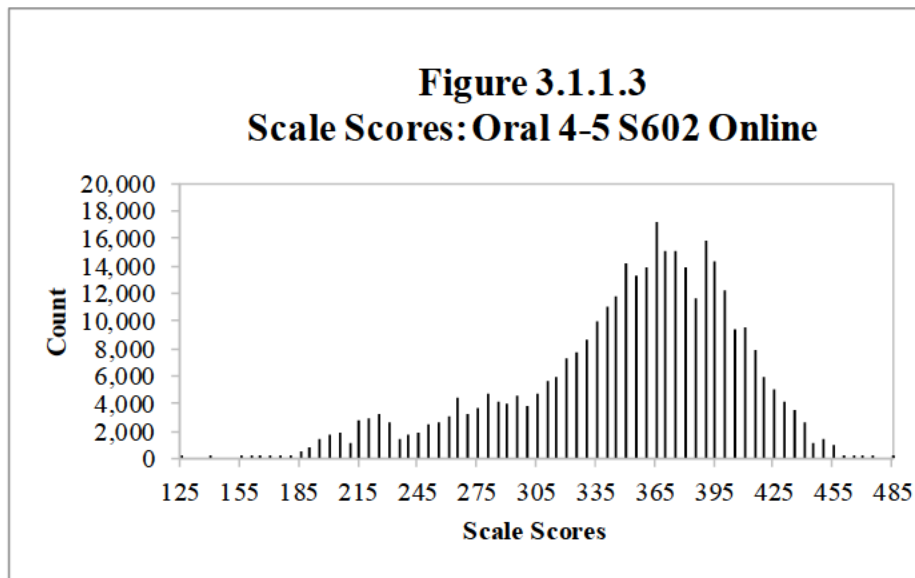
Table 3.1.1.3

Scale Score Descriptive Statistics: Oral 4–5 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	188,065	125	485	351.04	54.35
5	157,577	125	485	352.32	58.22
Total	345,642	125	485	351.63	56.15

Figure 3.1.1.3

Scale Scores: Oral 4–5 S602 Online



3.1.1.4 Grades 6–8

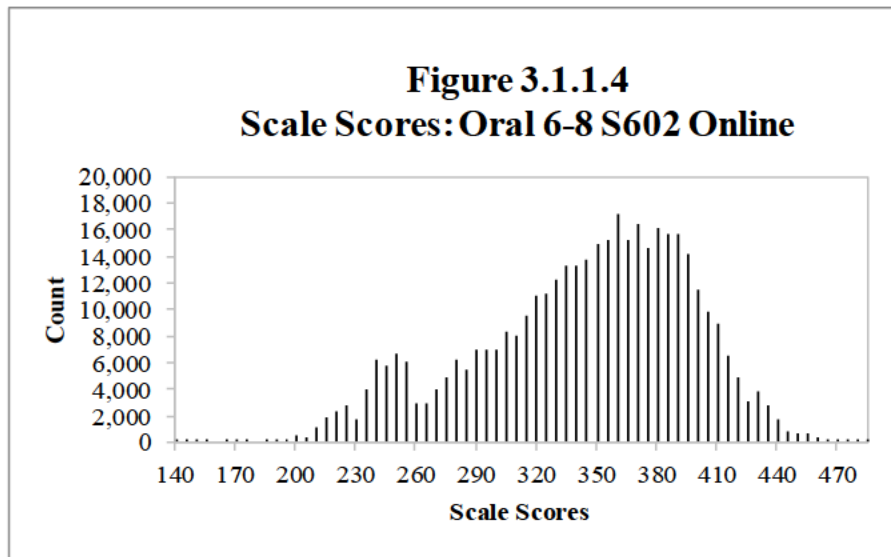
Table 3.1.1.4

Scale Score Descriptive Statistics: Oral 6–8 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	128,760	140	482	340.03	48.47
7	134,553	140	489	345.70	52.88
8	135,264	159	489	349.65	56.80
Total	398,577	140	489	345.21	53.04

Figure 3.1.1.4

Scale Scores: Oral 6–8 S602 Online



3.1.1.5 Grades 9–12

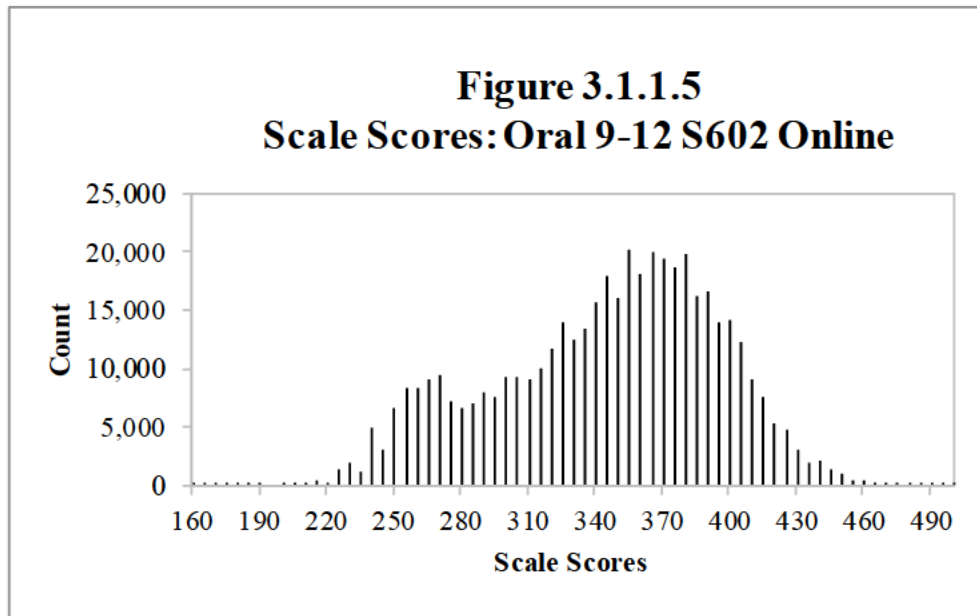
Table 3.1.1.5

Scale Score Descriptive Statistics: Oral 9–12 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	149,793	160	497	339.84	49.74
10	129,330	160	494	347.07	49.78
11	102,991	160	504	350.34	50.92
12	76,032	160	504	353.10	49.82
Total	458,146	160	504	346.44	50.28

Figure 3.1.1.5

Scale Scores: Oral 9–12 S602 Online



3.1.2 Literacy

3.1.2.1 Grade 1

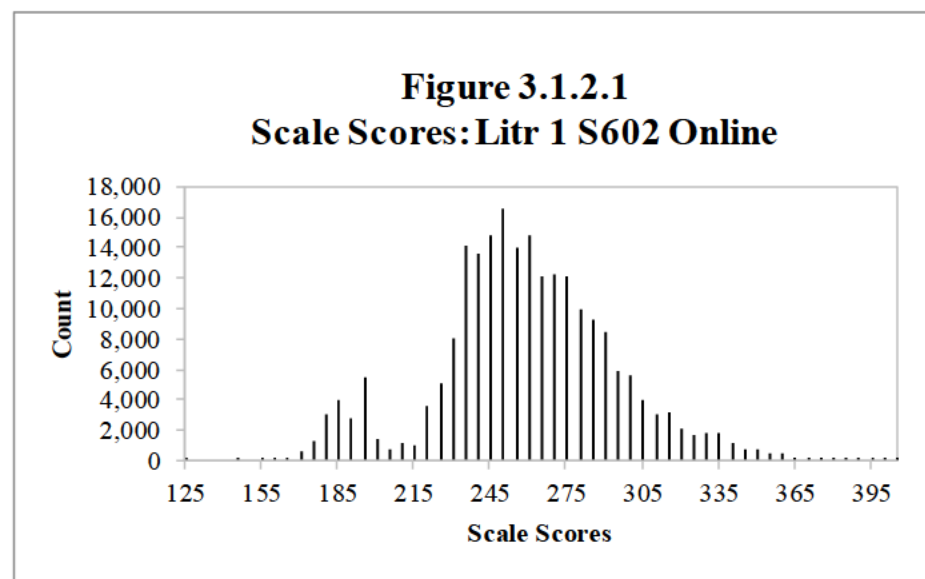
Table 3.1.2.1

Scale Score Descriptive Statistics: Litr 1 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	223,044	126	407	260.74	35.11
Total	223,044	126	407	260.74	35.11

Figure 3.1.2.1

Scale Scores: Litr 1 S602 Online



3.1.2.2 Grades 2–3

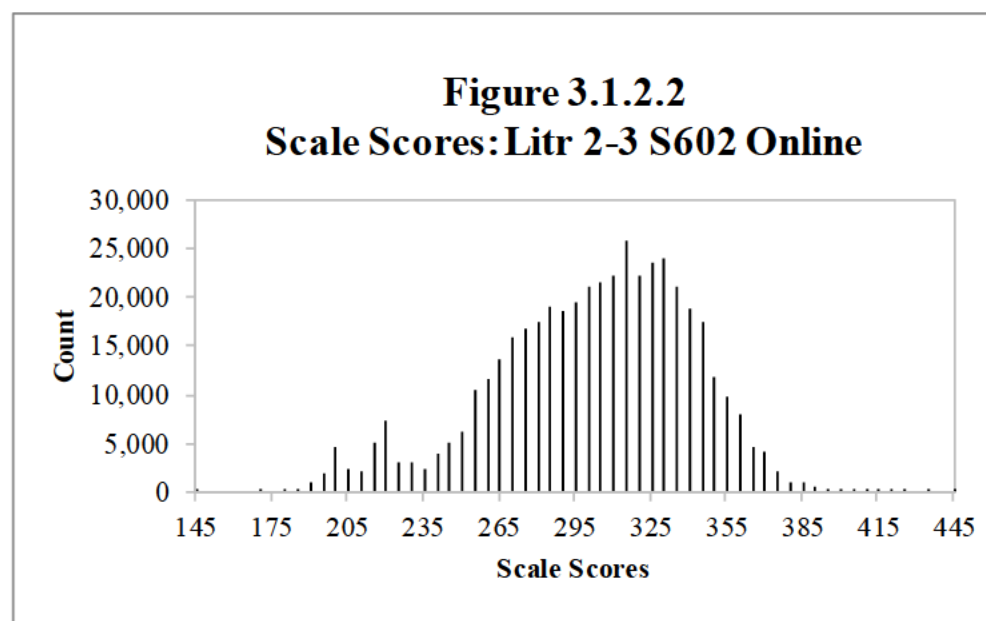
Table 3.1.2.2

Scale Score Descriptive Statistics: Litr 2–3 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	230,649	146	445	295.56	36.17
3	221,366	146	436	310.96	40.38
Total	452,015	146	445	303.10	39.06

Figure 3.1.2.2

Scale Scores: Litr 2–3 S602 Online



3.1.2.3 Grades 4–5

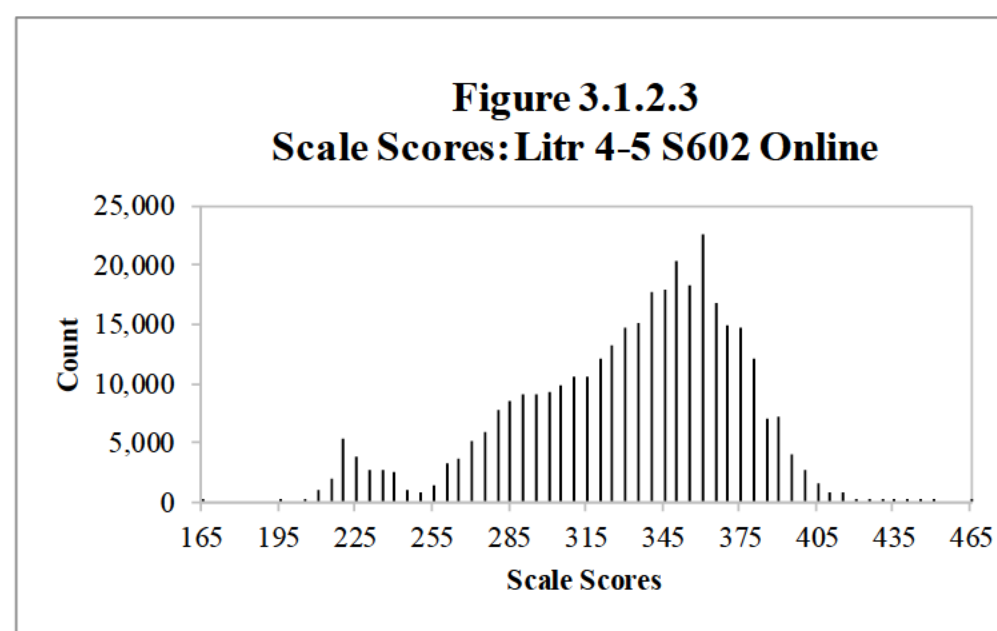
Table 3.1.2.3

Scale Score Descriptive Statistics: Litr 4–5 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	191,245	165	448	329.76	42.23
5	160,309	165	467	335.43	43.88
Total	351,554	165	467	332.35	43.09

Figure 3.1.2.3

Scale Scores: Litr 4–5 S602 Online



3.1.2.4 Grades 6–8

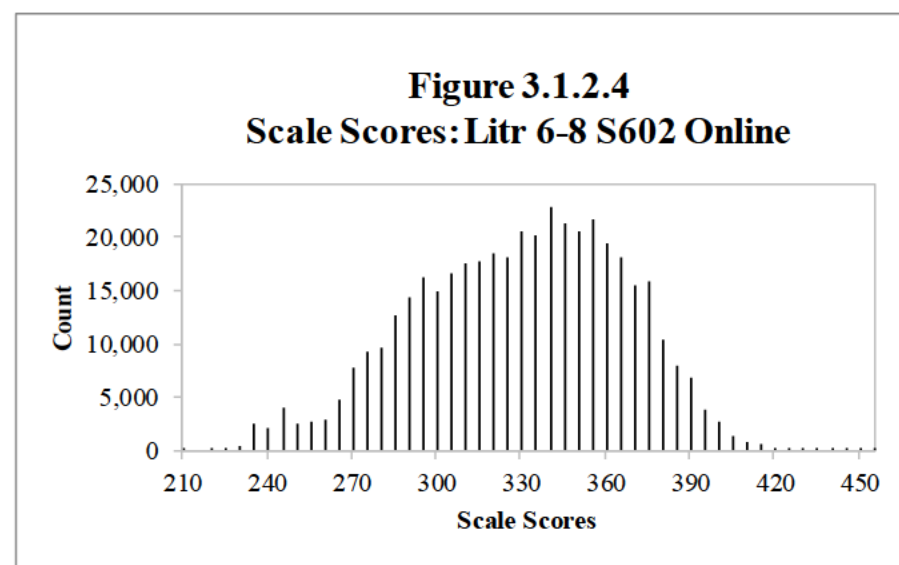
Table 3.1.2.4

Scale Score Descriptive Statistics: Litr 6–8 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	139,105	214	449	322.91	33.64
7	144,880	214	449	331.28	36.35
8	143,035	214	456	338.41	38.60
Total	427,020	214	456	330.94	36.82

Figure 3.1.2.4

Scale Scores: Litr 6–8 S602 Online



3.1.2.5 Grades 9–12

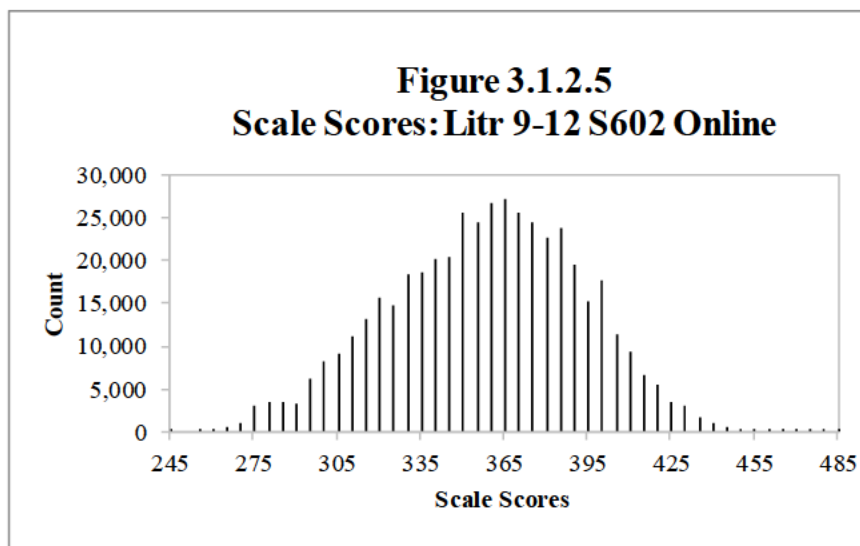
Table 3.1.2.5

Scale Score Descriptive Statistics: Litr 9–12 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	152,812	247	476	354.16	34.81
10	131,267	247	482	360.48	33.94
11	105,019	257	477	364.34	34.19
12	76,382	247	486	365.53	33.32
Total	465,480	247	486	360.11	34.48

Figure 3.1.2.5

Scale Scores: Litr 9–12 S602 Online



3.1.3 Comprehension

3.1.3.1 Grade 1

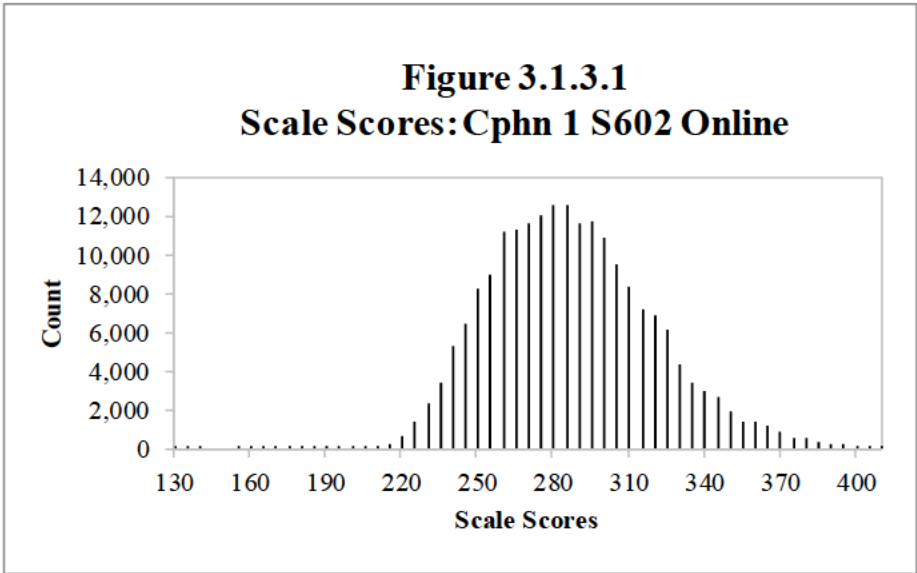
Table 3.1.3.1

Scale Score Descriptive Statistics: Cphn 1 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	203,660	130	412	289.23	32.28
Total	203,660	130	412	289.23	32.28

Figure 3.1.3.1

Scale Scores: Cphn 1 S602 Online



3.1.3.2 Grades 2–3

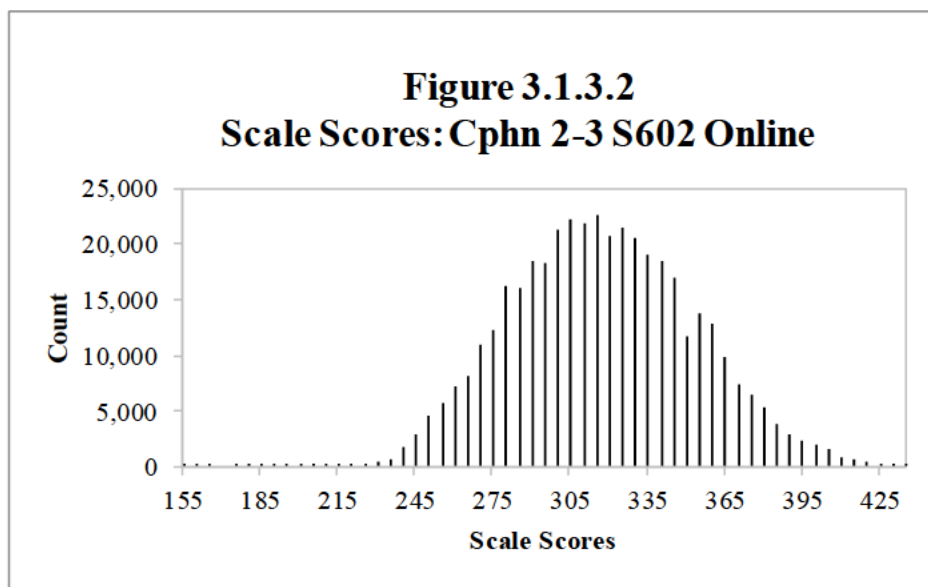
Table 3.1.3.2

Scale Score Descriptive Statistics: Cphn 2–3 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	208,640	161	439	312.79	30.73
3	203,385	155	439	326.01	38.40
Total	412,025	155	439	319.32	35.35

Figure 3.1.3.2

Scale Scores: Cphn 2–3 S602 Online



3.1.3.3 Grades 4–5

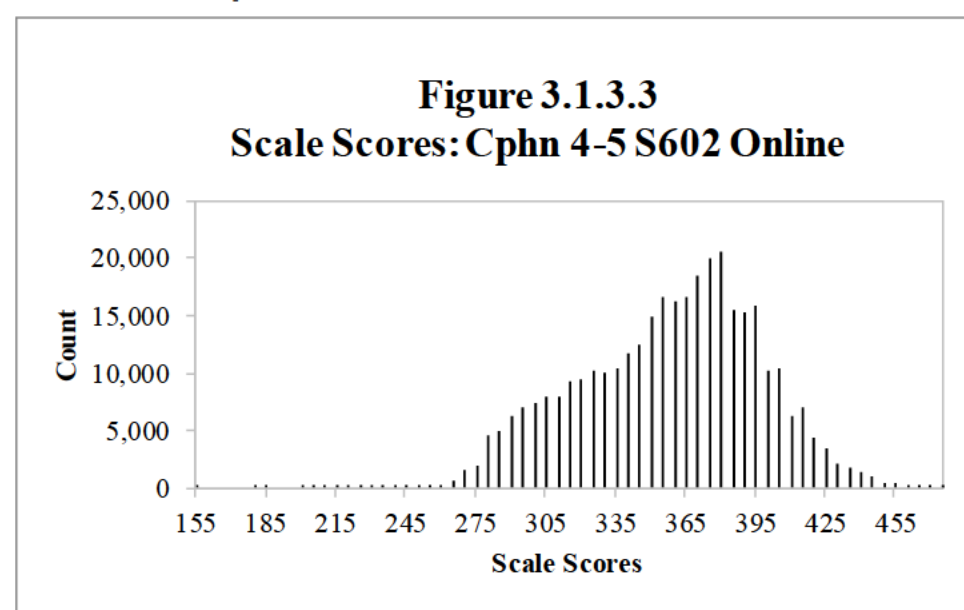
Table 3.1.3.3

Scale Score Descriptive Statistics: Cphn 4–5 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	187,289	159	475	358.44	37.02
5	156,746	159	475	361.83	39.85
Total	344,035	159	475	359.99	38.37

Figure 3.1.3.3

Scale Scores: Cphn 4–5 S602 Online



3.1.3.4 Grades 6–8

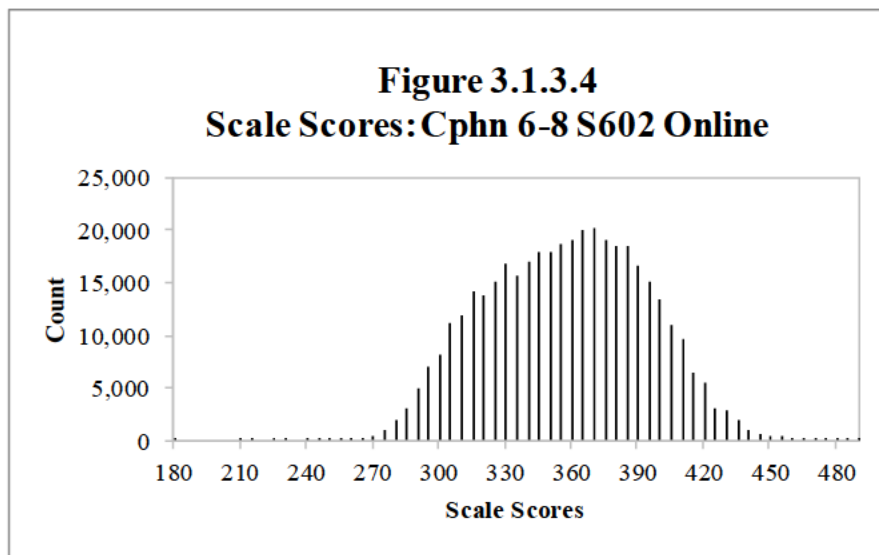
Table 3.1.3.4

Scale Score Descriptive Statistics: Cphn 6–8 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	129,782	211	492	352.18	32.07
7	135,867	180	492	359.65	35.76
8	135,675	231	492	365.84	38.83
Total	401,324	180	492	359.33	36.14

Figure 3.1.3.4

Scale Scores: Cphn 6–8 S602 Online



3.1.3.5 Grades 9–12

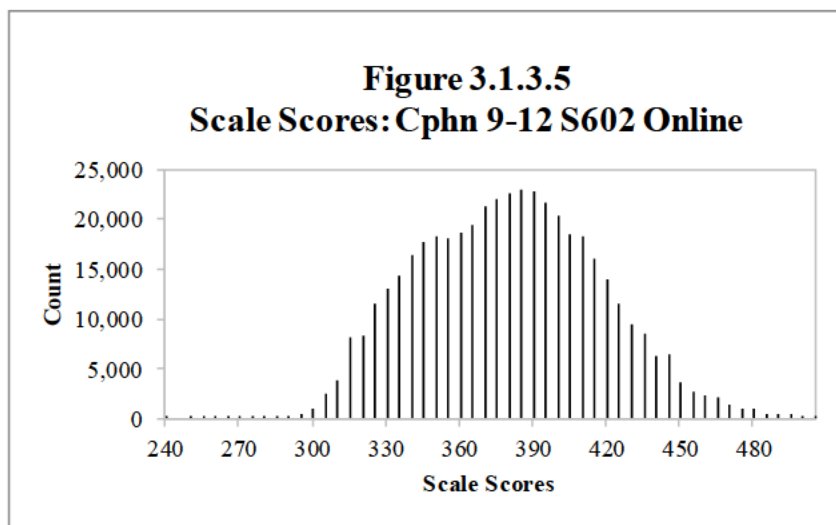
Table 3.1.3.5

Scale Score Descriptive Statistics: Cphn 9–12 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	147,401	257	508	376.40	35.36
10	127,547	262	508	382.33	36.64
11	102,031	241	508	385.36	38.38
12	74,372	267	508	387.02	38.00
Total	451,351	241	508	381.85	37.09

Figure 3.1.3.5

Scale Scores: Cphn 9–12 S602 Online



3.1.4 Overall

3.1.4.1 Grade 1

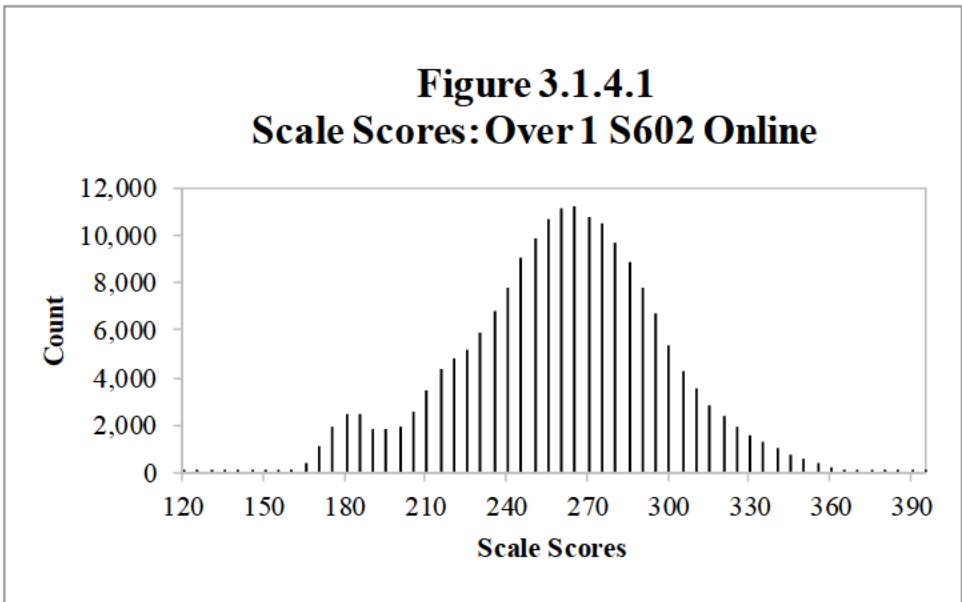
Table 3.1.4.1

Scale Score Descriptive Statistics: Over 1 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
1	187,784	120	399	261.91	37.09
Total	187,784	120	399	261.91	37.09

Figure 3.1.4.1

Scale Scores: Over 1 S602 Online



3.1.4.2 Grades 2–3

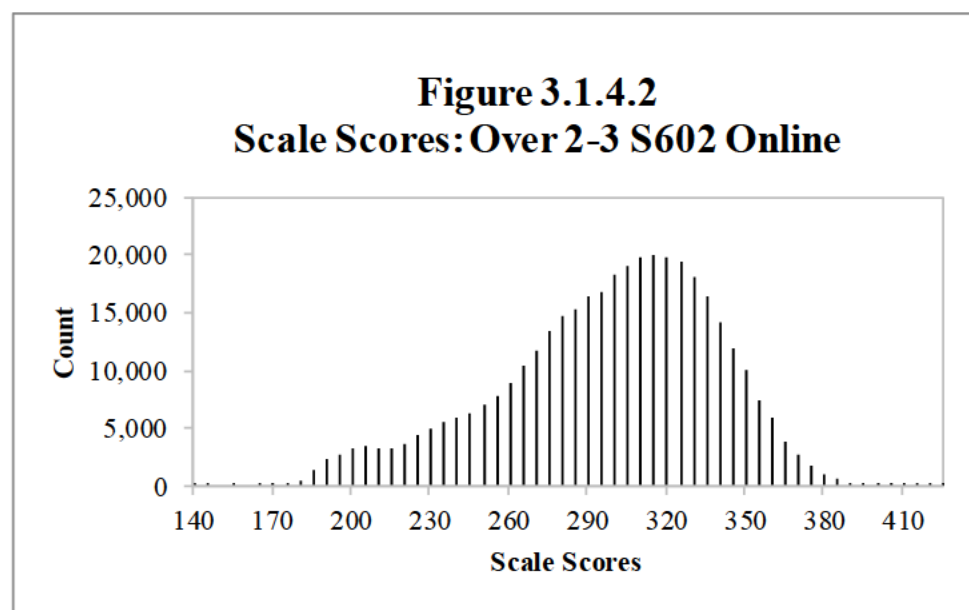
Table 3.1.4.2

Scale Score Descriptive Statistics: Over 2–3 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
2	193,331	145	427	291.01	37.86
3	190,836	142	423	307.00	42.91
Total	384,167	142	427	298.95	41.23

Figure 3.1.4.2

Scale Scores: Over 2–3 S602 Online



3.1.4.3 Grades 4–5

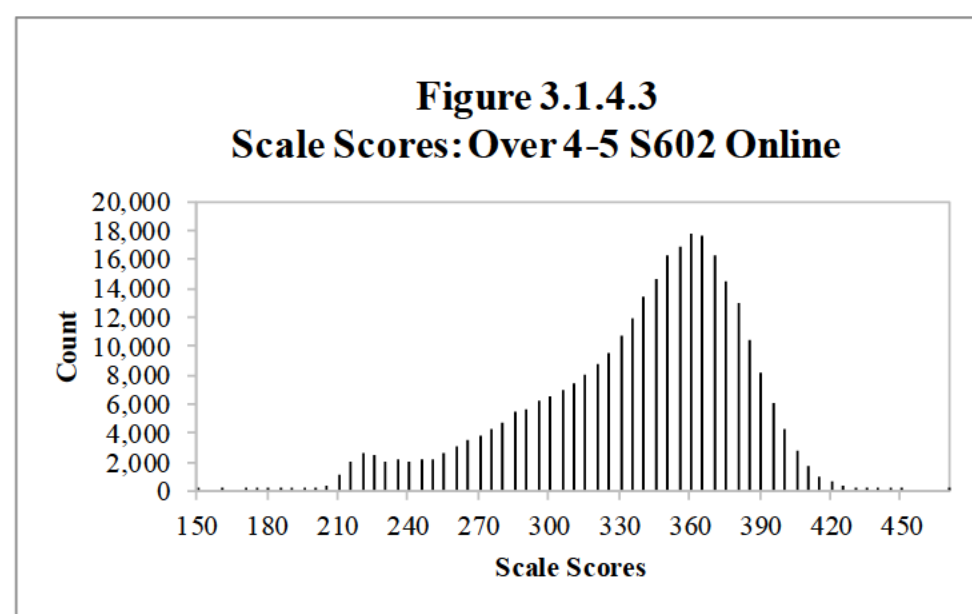
Table 3.1.4.3

Scale Score Descriptive Statistics: Over 4–5 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
4	164,672	164	450	336.15	43.44
5	139,738	153	472	340.39	45.94
Total	304,410	153	472	338.10	44.65

Figure 3.1.4.3

Scale Scores: Over 4–5 S602 Online



3.1.4.4 Grades 6–8

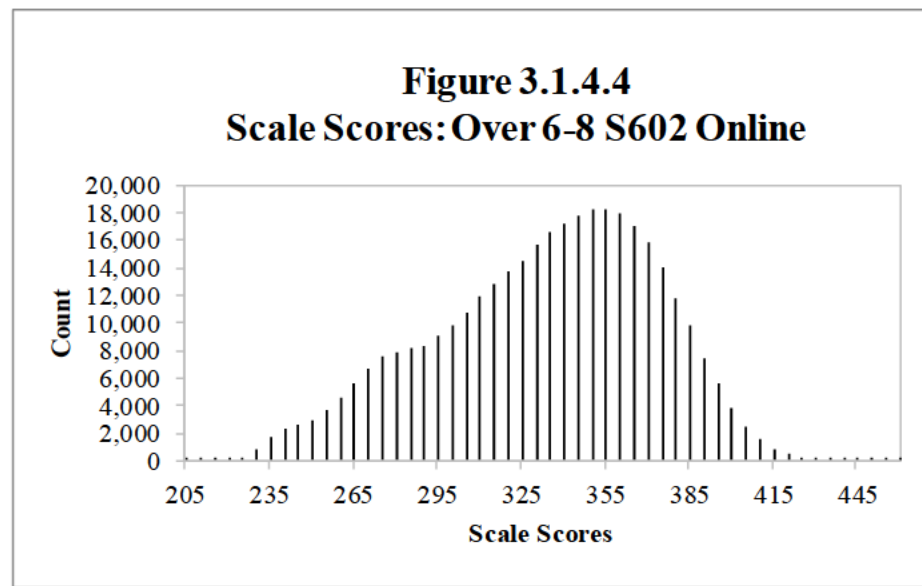
Table 3.1.4.4

Scale Score Descriptive Statistics: Over 6–8 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
6	115,623	209	451	328.07	35.78
7	121,006	211	459	335.48	39.22
8	121,717	209	462	341.36	42.08
Total	358,346	209	462	335.09	39.54

Figure 3.1.4.4

Scale Scores: Over 6–8 S602 Online



3.1.4.5 Grades 9–12

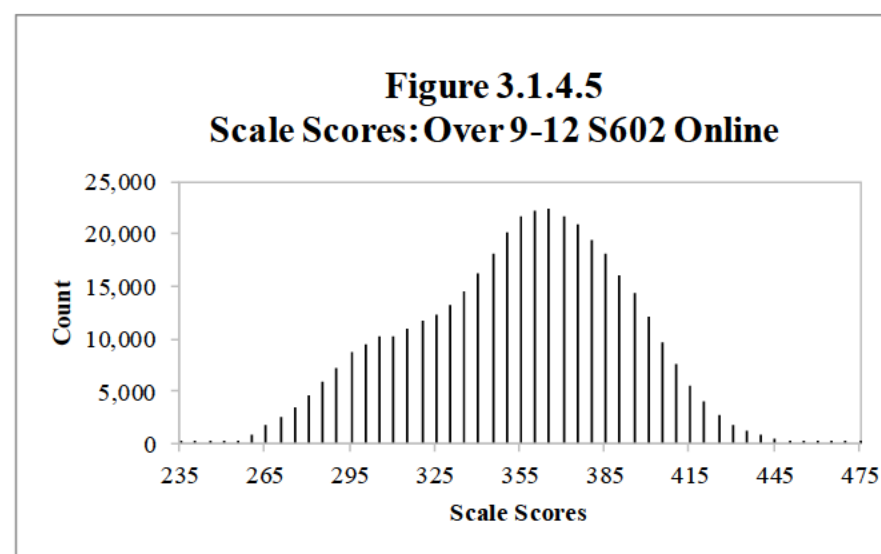
Table 3.1.4.5

Scale Score Descriptive Statistics: Over 9–12 S602 Online

Grade	# of Students	Min.	Max.	Mean	Std. Dev.
9	132,080	243	475	349.60	37.19
10	114,235	238	474	356.18	36.57
11	91,135	249	473	359.91	37.06
12	67,706	248	474	361.48	35.86
Total	405,156	238	475	355.76	37.06

Figure 3.1.4.5

Scale Scores: Over 9–12 S602 Online



3.2 Proficiency Level Distribution for Composites

Figures and tables in this section provide information on the proficiency level distribution for each of the composites for each grade-level cluster, denoted by *G#*.

In each figure, the horizontal axis shows the six WIDA proficiency levels. The vertical axis shows the percentage of students. Each bar shows the percentage of students placed into each proficiency level in the domain being tested on this test form.

The tables in this section present, by grade and by total for the grade-level cluster:

- The WIDA proficiency level designation (PL 1–PL 6)
- The number of students (count) whose performance on the test form placed them into that proficiency level in the domain being tested
- The percentage of students, out of the total number of students taking the form, who were placed into that proficiency level in the domain being tested

3.2.1 Oral

3.2.1.1 Grade 1

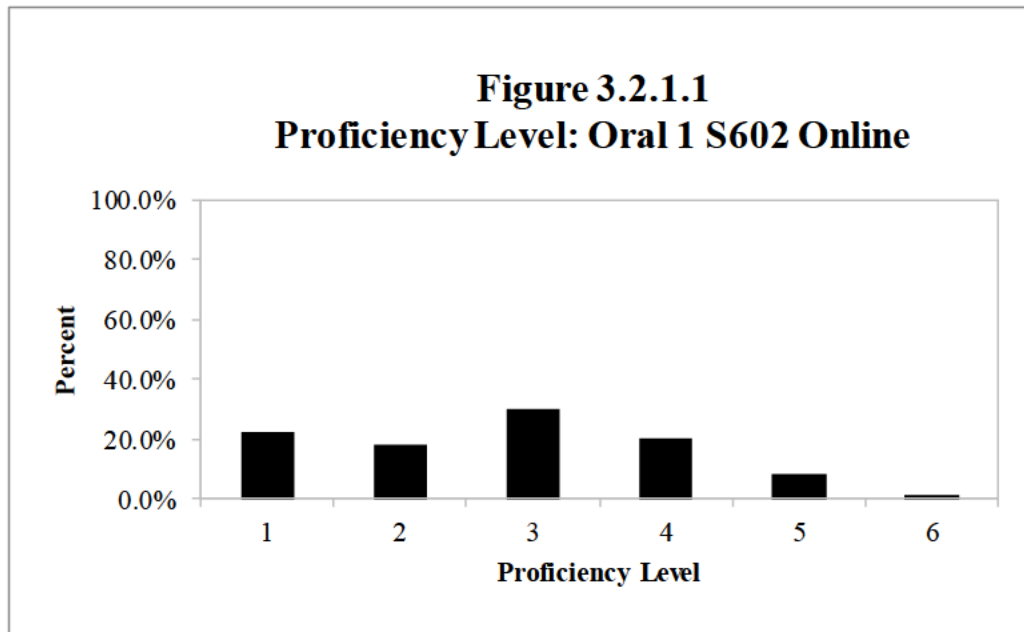
Table 3.2.1.1

Proficiency Level Distribution: Oral 1 S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	44,045	22.46%	44,045	22.46%
2	35,361	18.03%	35,361	18.03%
3	59,286	30.23%	59,286	30.23%
4	39,637	20.21%	39,637	20.21%
5	15,951	8.13%	15,951	8.13%
6	1,865	0.95%	1,865	0.95%
Total	196,145	100.00%	196,145	100.00%

Figure 3.2.1.1

Proficiency Level: Oral 1 S602 Online



3.2.1.2 Grades 2–3

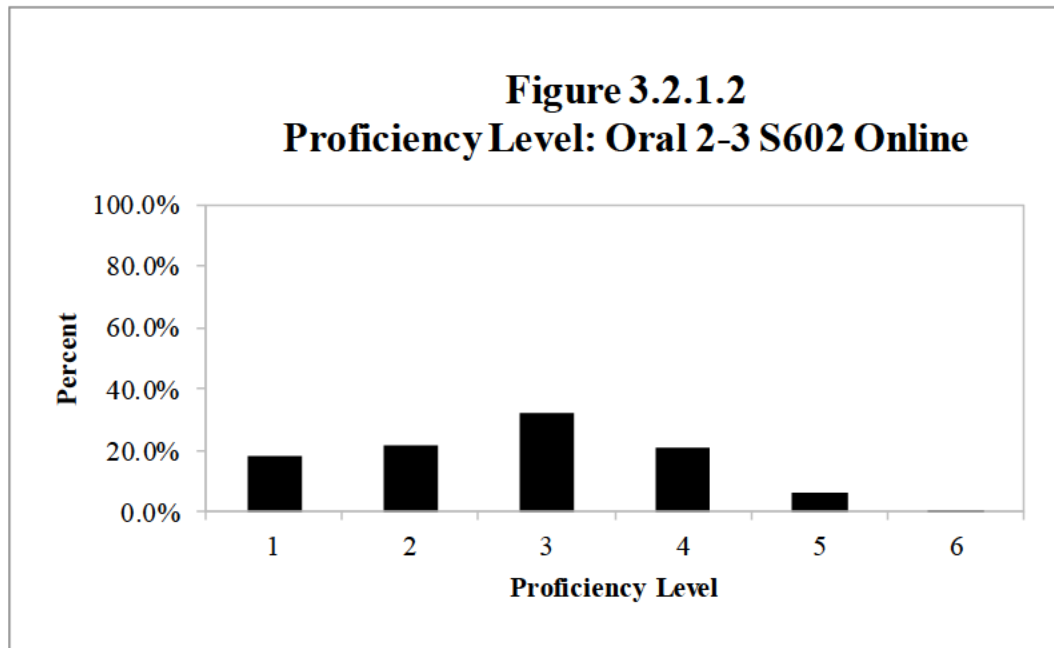
Table 3.2.1.2

Proficiency Level Distribution: Oral 2–3 S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	36,328	17.80%	36,431	18.08%	72,759	17.94%
2	49,528	24.27%	38,043	18.88%	87,571	21.59%
3	66,896	32.78%	64,597	32.06%	131,493	32.42%
4	38,371	18.80%	46,435	23.04%	84,806	20.91%
5	11,777	5.77%	14,330	7.11%	26,107	6.44%
6	1,184	0.58%	1,679	0.83%	2,863	0.71%
Total	204,084	100.00%	201,515	100.00%	405,599	100.00%

Figure 3.2.1.2

Proficiency Level: Oral 2–3 S602 Online



3.2.1.3 Grades 4–5

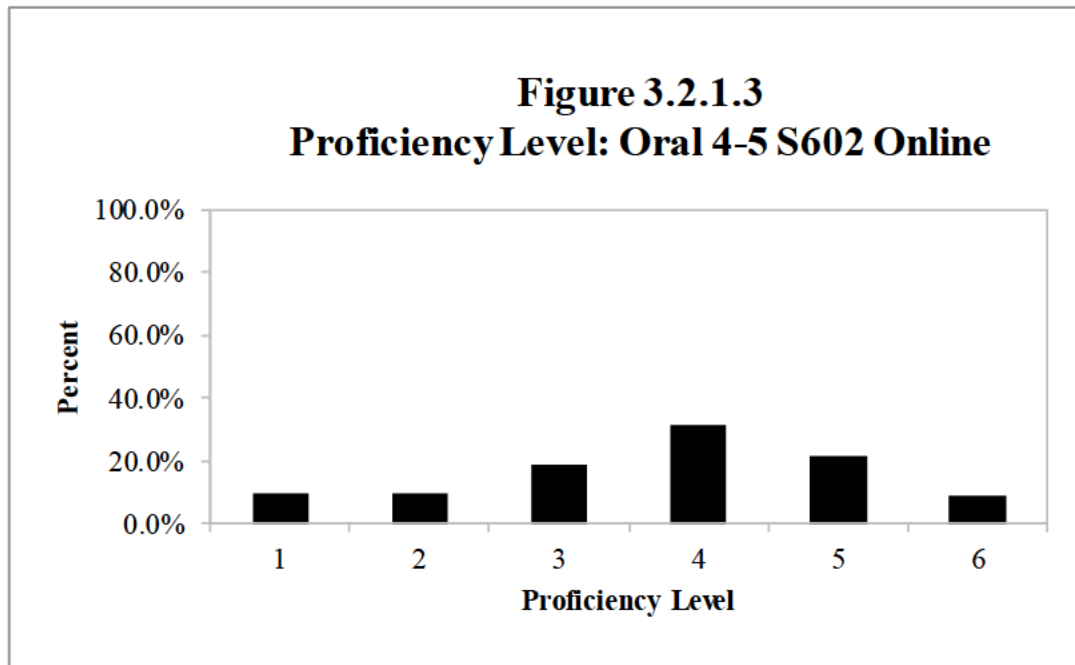
Table 3.2.1.3

Proficiency Level Distribution: Oral 4–5 S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	14,523	7.72%	18,459	11.71%	32,982	9.54%
2	19,408	10.32%	14,839	9.42%	34,247	9.91%
3	34,756	18.48%	30,122	19.12%	64,878	18.77%
4	57,100	30.36%	50,879	32.29%	107,979	31.24%
5	42,439	22.57%	31,661	20.09%	74,100	21.44%
6	19,839	10.55%	11,617	7.37%	31,456	9.10%
Total	188,065	100.00%	157,577	100.00%	345,642	100.00%

Figure 3.2.1.3

Proficiency Level: Oral 4–5 S602 Online



3.2.1.4 Grades 6–8

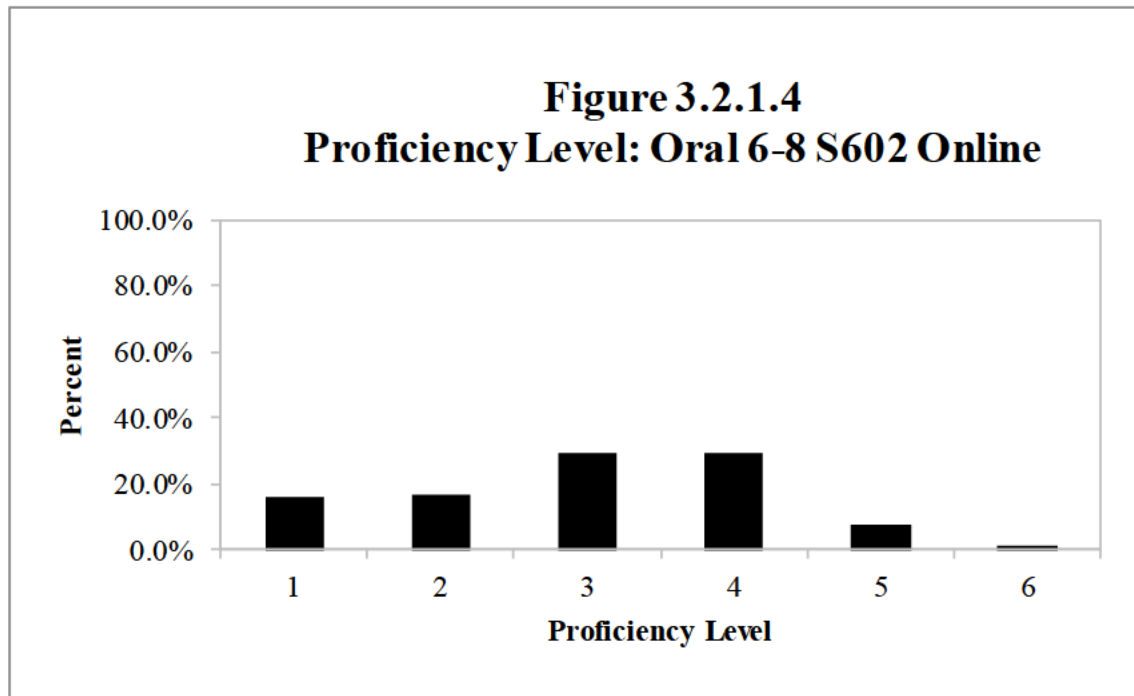
Table 3.2.1.4

Proficiency Level Distribution: Oral 6–8 S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	16,657	12.94%	21,843	16.23%	25,700	19.00%	64,200	16.11%
2	22,922	17.80%	22,354	16.61%	20,780	15.36%	66,056	16.57%
3	41,201	32.00%	39,454	29.32%	37,468	27.70%	118,123	29.64%
4	37,845	29.39%	39,531	29.38%	39,074	28.89%	116,450	29.22%
5	9,020	7.01%	10,092	7.50%	10,736	7.94%	29,848	7.49%
6	1,115	0.87%	1,279	0.95%	1,506	1.11%	3,900	0.98%
Total	128,760	100.00%	134,553	100.00%	135,264	100.00%	398,577	100.00%

Figure 3.2.1.4

Proficiency Level: Oral 6–8 S602 Online



3.2.1.5 Grades 9–12

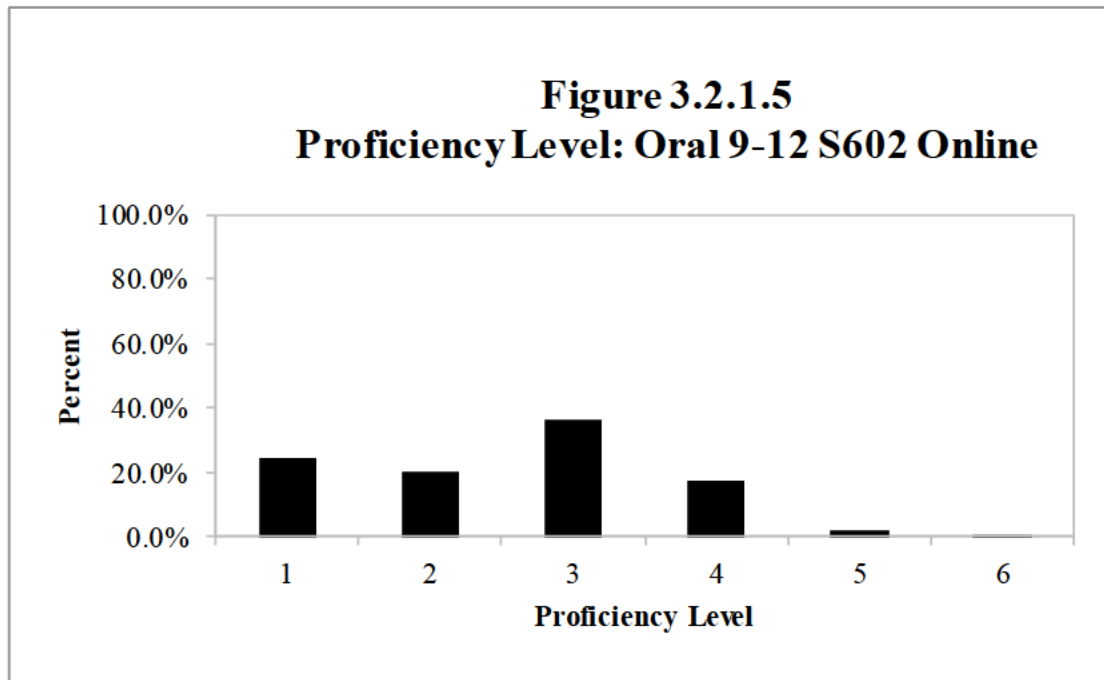
Table 3.2.1.5

Proficiency Level Distribution: Oral 9–12 S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	36,087	24.09%	30,450	23.54%	25,695	24.95%	18,808	24.74%	111,040	24.24%
2	30,700	20.49%	26,047	20.14%	20,488	19.89%	14,912	19.61%	92,147	20.11%
3	52,673	35.16%	46,551	35.99%	37,225	36.14%	29,947	39.39%	166,396	36.32%
4	27,325	18.24%	23,411	18.10%	17,160	16.66%	11,039	14.52%	78,935	17.23%
5	2,659	1.78%	2,563	1.98%	2,154	2.09%	1,173	1.54%	8,549	1.87%
6	349	0.23%	308	0.24%	269	0.26%	153	0.20%	1,079	0.24%
Total	149,793	100.00%	129,330	100.00%	102,991	100.00%	76,032	100.00%	458,146	100.00%

Figure 3.2.1.5

Proficiency Level: Oral 9–12 S602 Online



3.2.2 Literacy

3.2.2.1 Grade 1

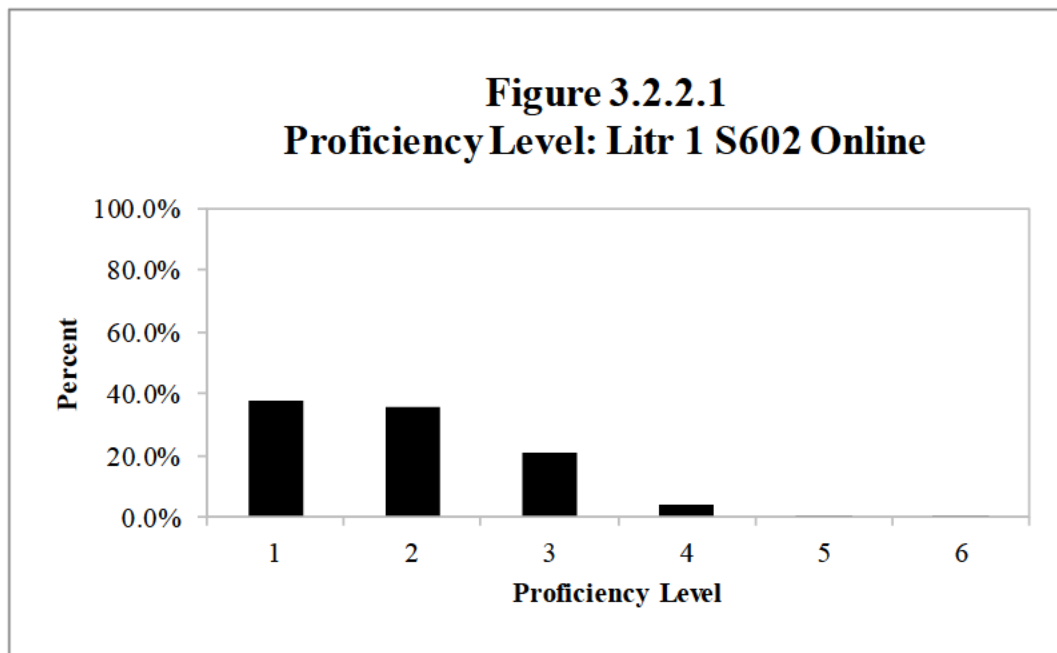
Table 3.2.2.1

Proficiency Level Distribution: Litr 1 S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	84,731	37.99%	84,731	37.99%
2	79,732	35.75%	79,732	35.75%
3	47,278	21.20%	47,278	21.20%
4	9,044	4.05%	9,044	4.05%
5	1,893	0.85%	1,893	0.85%
6	366	0.16%	366	0.16%
Total	223,044	100.00%	223,044	100.00%

Figure 3.2.2.1

Proficiency Level: Litr 1 S602 Online



3.2.2.2 Grades 2–3

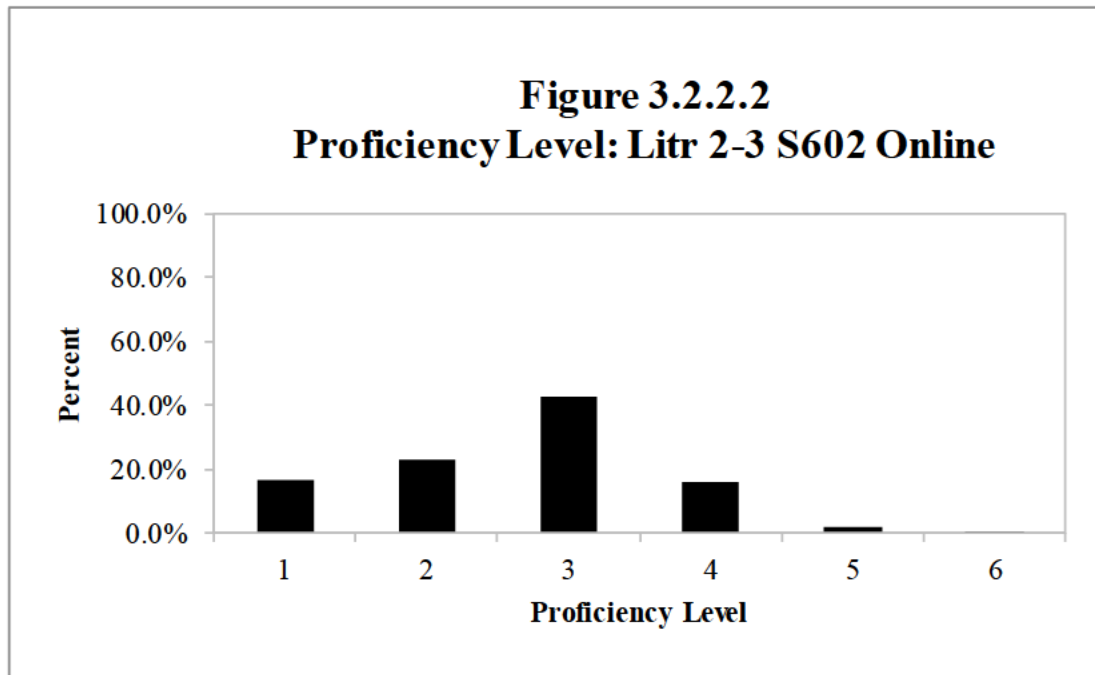
Table 3.2.2.2

Proficiency Level Distribution: Litr 2–3 S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	37,621	16.31%	36,769	16.61%	74,390	16.46%
2	60,850	26.38%	43,621	19.71%	104,471	23.11%
3	99,894	43.31%	93,050	42.03%	192,944	42.69%
4	29,800	12.92%	42,162	19.05%	71,962	15.92%
5	2,222	0.96%	5,284	2.39%	7,506	1.66%
6	262	0.11%	480	0.22%	742	0.16%
Total	230,649	100.00%	221,366	100.00%	452,015	100.00%

Figure 3.2.2.2

Proficiency Level: Litr 2–3 S602 Online



3.2.2.3 Grades 4–5

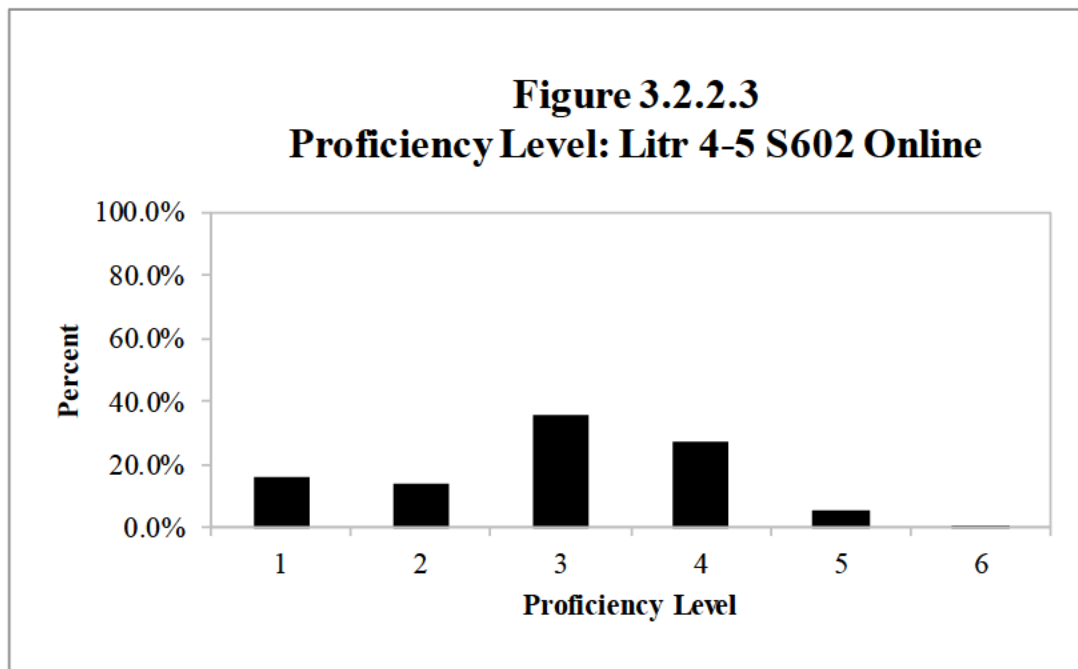
Table 3.2.2.3

Proficiency Level Distribution: Litr 4–5 S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	29,675	15.52%	27,016	16.85%	56,691	16.13%
2	26,263	13.73%	22,191	13.84%	48,454	13.78%
3	70,745	36.99%	55,713	34.75%	126,458	35.97%
4	52,042	27.21%	44,870	27.99%	96,912	27.57%
5	10,709	5.60%	9,346	5.83%	20,055	5.70%
6	1,811	0.95%	1,173	0.73%	2,984	0.85%
Total	191,245	100.00%	160,309	100.00%	351,554	100.00%

Figure 3.2.2.3

Proficiency Level: Litr 4–5 S602 Online



3.2.2.4 Grades 6–8

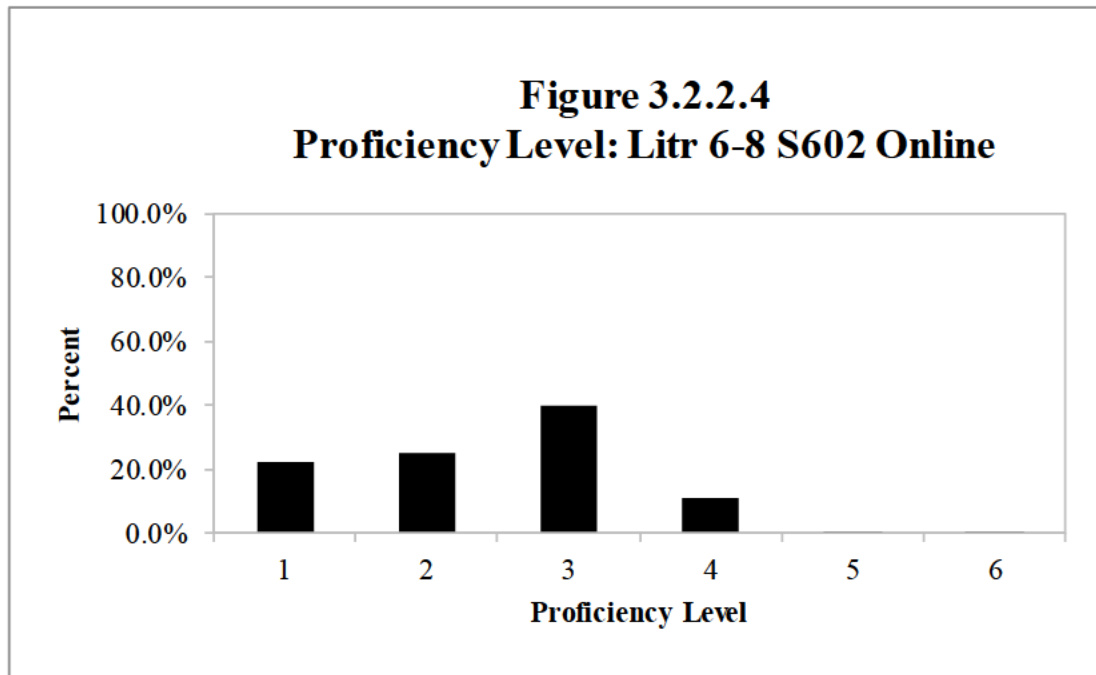
Table 3.2.2.4

Proficiency Level Distribution: Litr 6–8 S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	31,360	22.54%	31,205	21.54%	33,940	23.73%	96,505	22.60%
2	37,552	27.00%	37,883	26.15%	32,126	22.46%	107,561	25.19%
3	58,437	42.01%	58,646	40.48%	54,554	38.14%	171,637	40.19%
4	11,215	8.06%	16,110	11.12%	21,040	14.71%	48,365	11.33%
5	505	0.36%	981	0.68%	1,329	0.93%	2,815	0.66%
6	36	0.03%	55	0.04%	46	0.03%	137	0.03%
Total	139,105	100.00%	144,880	100.00%	143,035	100.00%	427,020	100.00%

Figure 3.2.2.4

Proficiency Level: Litr 6–8 S602 Online



3.2.2.5 Grades 9–12

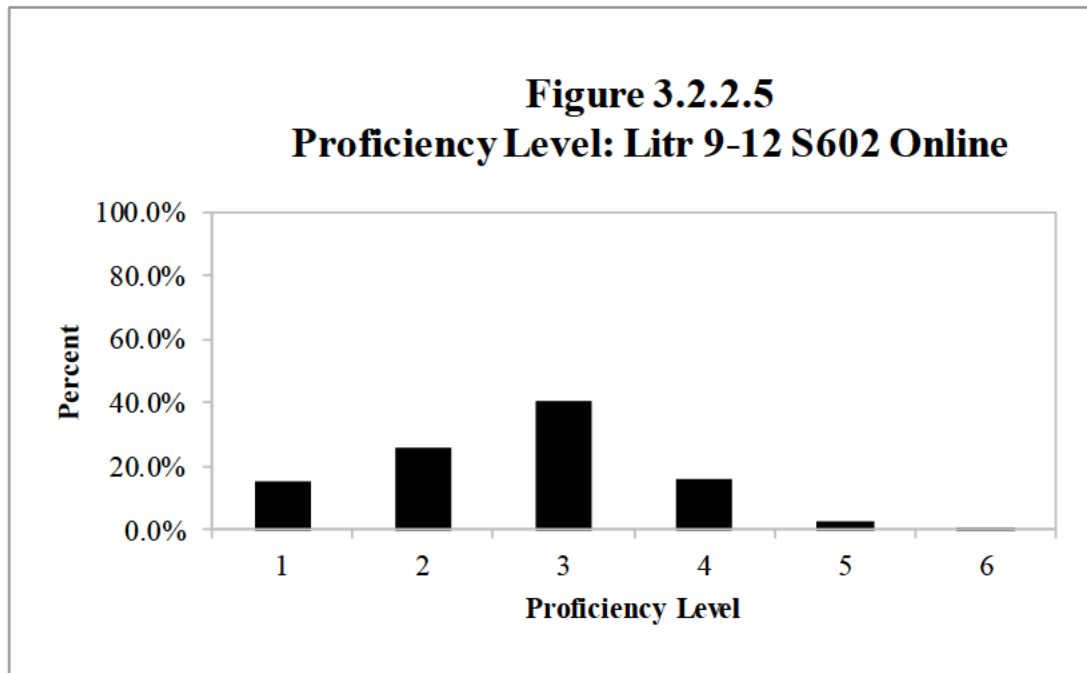
Table 3.2.2.5

Proficiency Level Distribution: Litr 9–12 S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	22,505	14.73%	17,233	13.13%	16,267	15.49%	13,864	18.15%	69,869	15.01%
2	35,986	23.55%	33,310	25.38%	27,675	26.35%	22,457	29.40%	119,428	25.66%
3	63,197	41.36%	54,407	41.45%	42,070	40.06%	29,787	39.00%	189,461	40.70%
4	26,793	17.53%	22,556	17.18%	16,017	15.25%	9,004	11.79%	74,370	15.98%
5	4,060	2.66%	3,595	2.74%	2,927	2.79%	1,259	1.65%	11,841	2.54%
6	271	0.18%	166	0.13%	63	0.06%	11	0.01%	511	0.11%
Total	152,812	100.00%	131,267	100.00%	105,019	100.00%	76,382	100.00%	465,480	100.00%

Figure 3.2.2.5

Proficiency Level: Litr 9–12 S602 Online



3.2.3 Comprehension

3.2.3.1 Grade 1

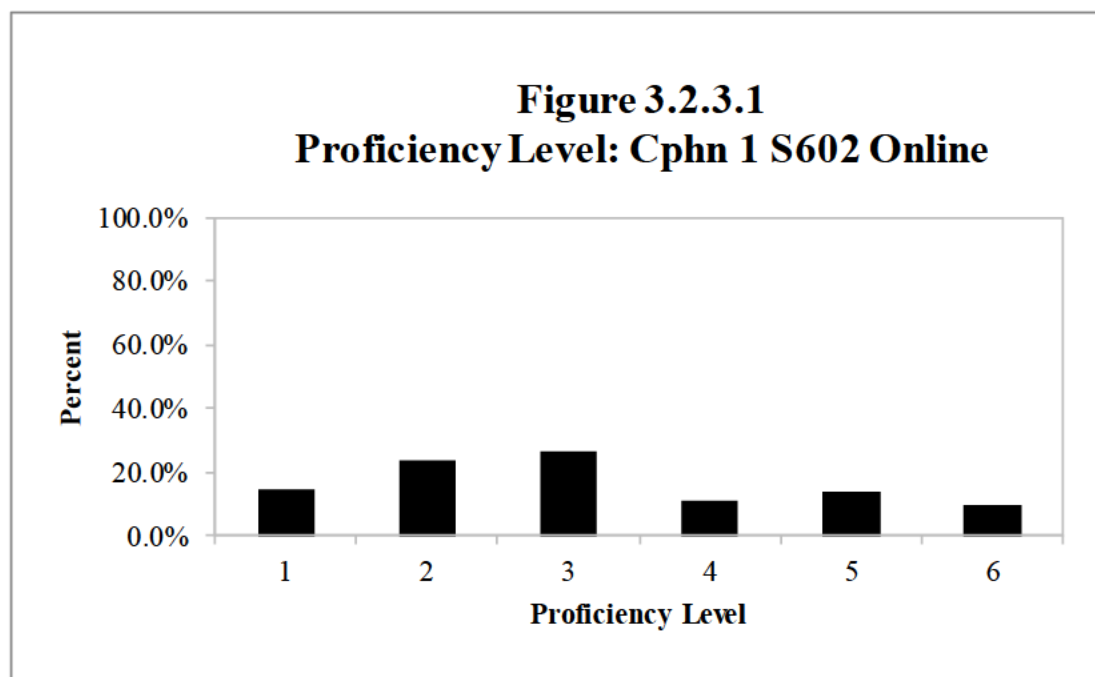
Table 3.2.3.1

Proficiency Level Distribution: Cphn 1 S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	30,073	14.77%	30,073	14.77%
2	48,399	23.76%	48,399	23.76%
3	53,740	26.39%	53,740	26.39%
4	22,295	10.95%	22,295	10.95%
5	28,737	14.11%	28,737	14.11%
6	20,416	10.02%	20,416	10.02%
Total	203,660	100.00%	203,660	100.00%

Figure 3.2.3.1

Proficiency Level: Cphn 1 S602 Online



3.2.3.2 Grades 2–3

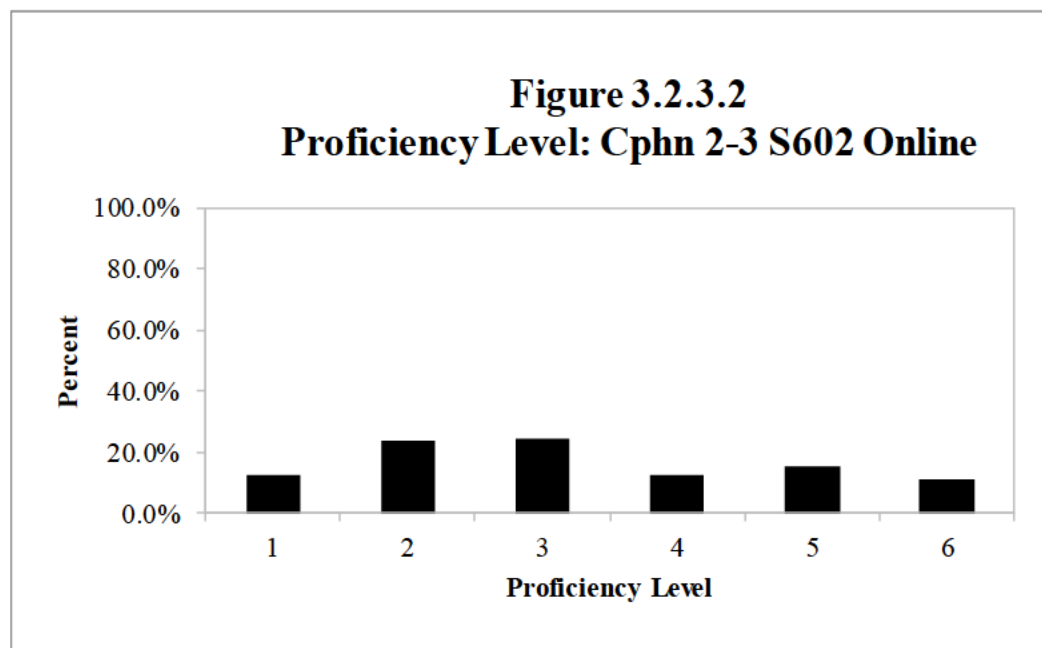
Table 3.2.3.2

Proficiency Level Distribution: Cphn 2–3 S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	17,851	8.56%	34,986	17.20%	52,837	12.82%
2	52,591	25.21%	45,177	22.21%	97,768	23.73%
3	57,988	27.79%	44,081	21.67%	102,069	24.77%
4	30,210	14.48%	21,865	10.75%	52,075	12.64%
5	30,495	14.62%	31,543	15.51%	62,038	15.06%
6	19,505	9.35%	25,733	12.65%	45,238	10.98%
Total	208,640	100.00%	203,385	100.00%	412,025	100.00%

Figure 3.2.3.2

Proficiency Level: Cphn 2–3 S602 Online



3.2.3.3 Grades 4–5

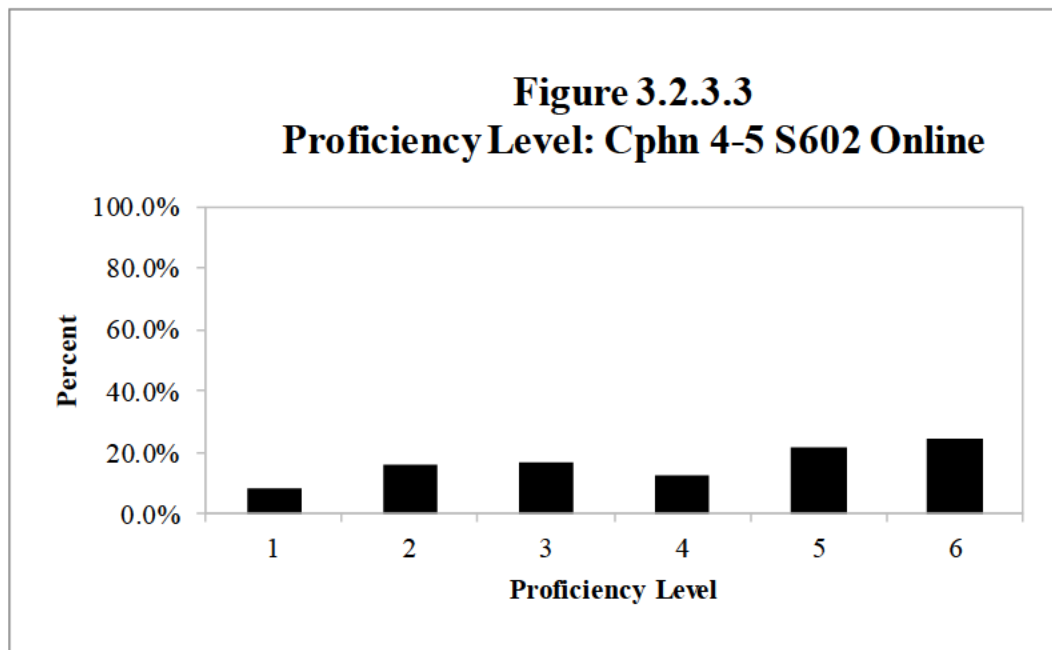
Table 3.2.3.3

Proficiency Level Distribution: Cphn 4–5 S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	11,149	5.95%	18,665	11.91%	29,814	8.67%
2	30,728	16.41%	23,834	15.21%	54,562	15.86%
3	31,005	16.55%	25,426	16.22%	56,431	16.40%
4	22,893	12.22%	20,760	13.24%	43,653	12.69%
5	40,948	21.86%	33,562	21.41%	74,510	21.66%
6	50,566	27.00%	34,499	22.01%	85,065	24.73%
Total	187,289	100.00%	156,746	100.00%	344,035	100.00%

Figure 3.2.3.3

Proficiency Level: Cphn 4–5 S602 Online



3.2.3.4 Grades 6-8

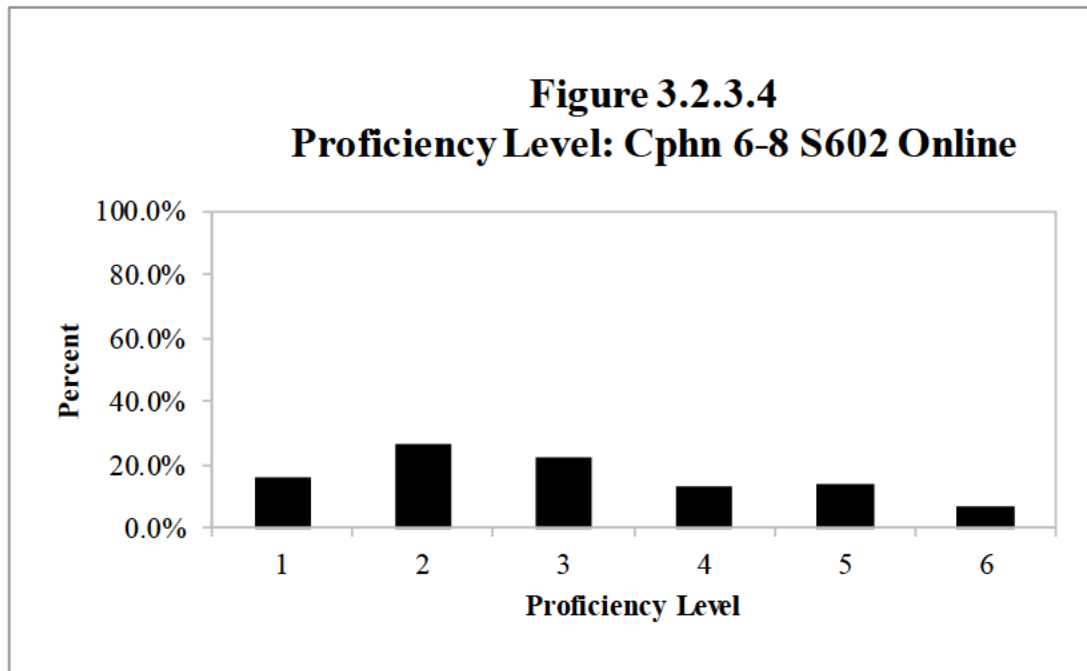
Table 3.2.3.4

Proficiency Level Distribution: Cphn 6-8 S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	16,560	12.76%	22,012	16.20%	25,776	19.00%	64,348	16.03%
2	39,593	30.51%	36,008	26.50%	31,442	23.17%	107,043	26.67%
3	33,125	25.52%	29,870	21.98%	27,153	20.01%	90,148	22.46%
4	16,736	12.90%	18,731	13.79%	18,422	13.58%	53,889	13.43%
5	16,875	13.00%	18,832	13.86%	21,120	15.57%	56,827	14.16%
6	6,893	5.31%	10,414	7.66%	11,762	8.67%	29,069	7.24%
Total	129,782	100.00%	135,867	100.00%	135,675	100.00%	401,324	100.00%

Figure 3.2.3.4

Proficiency Level: Cphn 6-8 S602 Online



3.2.3.5 Grades 9–12

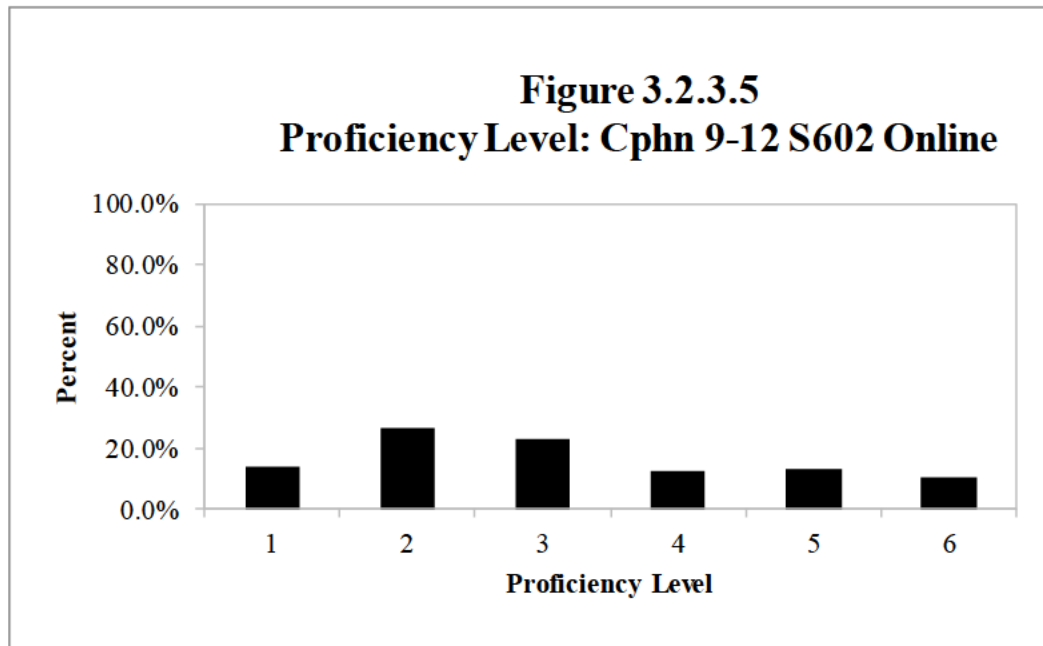
Table 3.2.3.5

Proficiency Level Distribution: Cphn 9–12 S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	16,462	11.17%	15,692	12.30%	16,347	16.02%	13,004	17.49%	61,505	13.63%
2	40,734	27.63%	33,267	26.08%	26,230	25.71%	19,942	26.81%	120,173	26.63%
3	35,239	23.91%	30,633	24.02%	22,417	21.97%	16,470	22.15%	104,759	23.21%
4	19,919	13.51%	16,674	13.07%	11,590	11.36%	8,811	11.85%	56,994	12.63%
5	20,735	14.07%	17,117	13.42%	13,506	13.24%	8,826	11.87%	60,184	13.33%
6	14,312	9.71%	14,164	11.10%	11,941	11.70%	7,319	9.84%	47,736	10.58%
Total	147,401	100.00%	127,547	100.00%	102,031	100.00%	74,372	100.00%	451,351	100.00%

Figure 3.2.3.5

Proficiency Level: Cphn 9–12 S602 Online



3.2.4 Overall

3.2.4.1 Grade 1

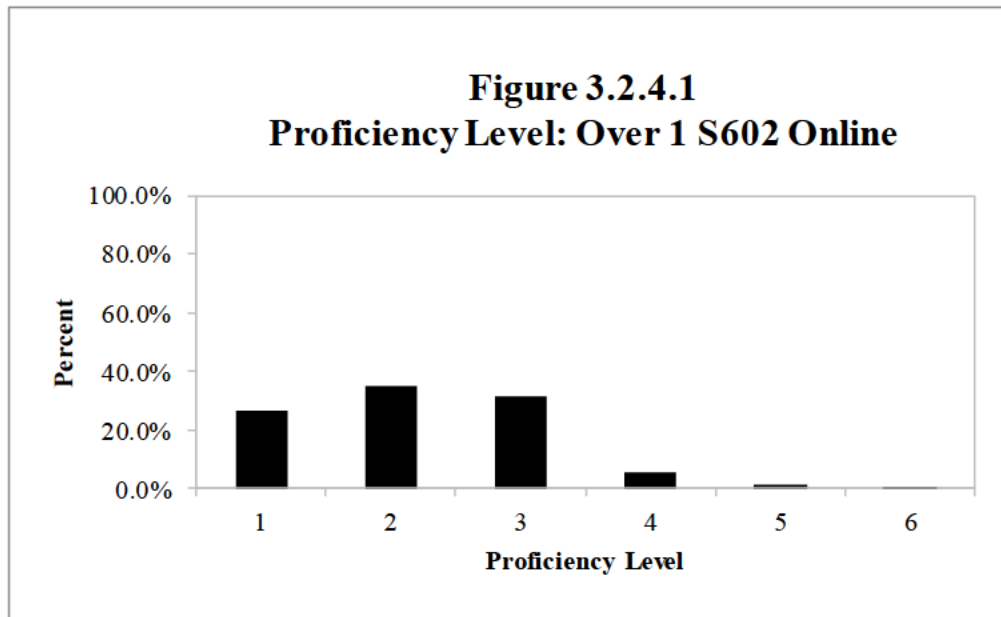
Table 3.2.4.1

Proficiency Level Distribution: Over 1 S602 Online

Level	G1 Count	G1 Percent	Total Count	Total Percent
1	50,185	26.72%	50,185	26.72%
2	65,500	34.88%	65,500	34.88%
3	58,706	31.26%	58,706	31.26%
4	10,889	5.80%	10,889	5.80%
5	2,274	1.21%	2,274	1.21%
6	230	0.12%	230	0.12%
Total	187,784	100.00%	187,784	100.00%

Figure 3.2.4.1

Proficiency Level: Over 1 S602 Online



3.2.4.2 Grades 2–3

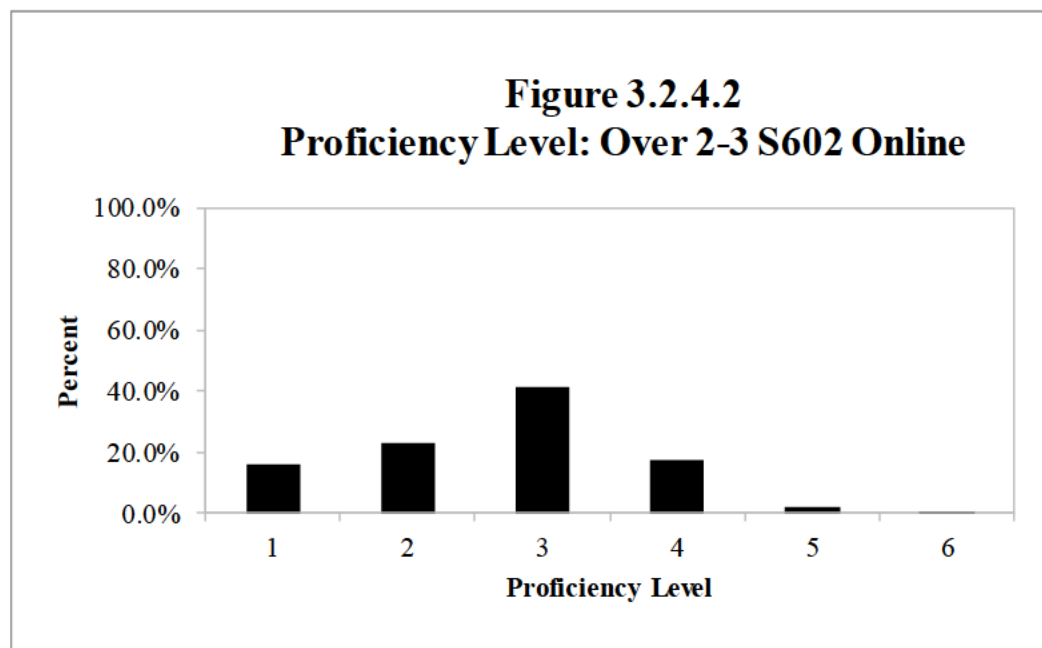
Table 3.2.4.2

Proficiency Level Distribution: Over 2–3 S602 Online

Level	G2 Count	G2 Percent	G3 Count	G3 Percent	Total Count	Total Percent
1	31,170	16.12%	31,787	16.66%	62,957	16.39%
2	51,162	26.46%	37,123	19.45%	88,285	22.98%
3	81,153	41.98%	76,705	40.19%	157,858	41.09%
4	27,005	13.97%	40,046	20.98%	67,051	17.45%
5	2,708	1.40%	4,985	2.61%	7,693	2.00%
6	133	0.07%	190	0.10%	323	0.08%
Total	193,331	100.00%	190,836	100.00%	384,167	100.00%

Figure 3.2.4.2

Proficiency Level: Over 2–3 S602 Online



3.2.4.3 Grades 4–5

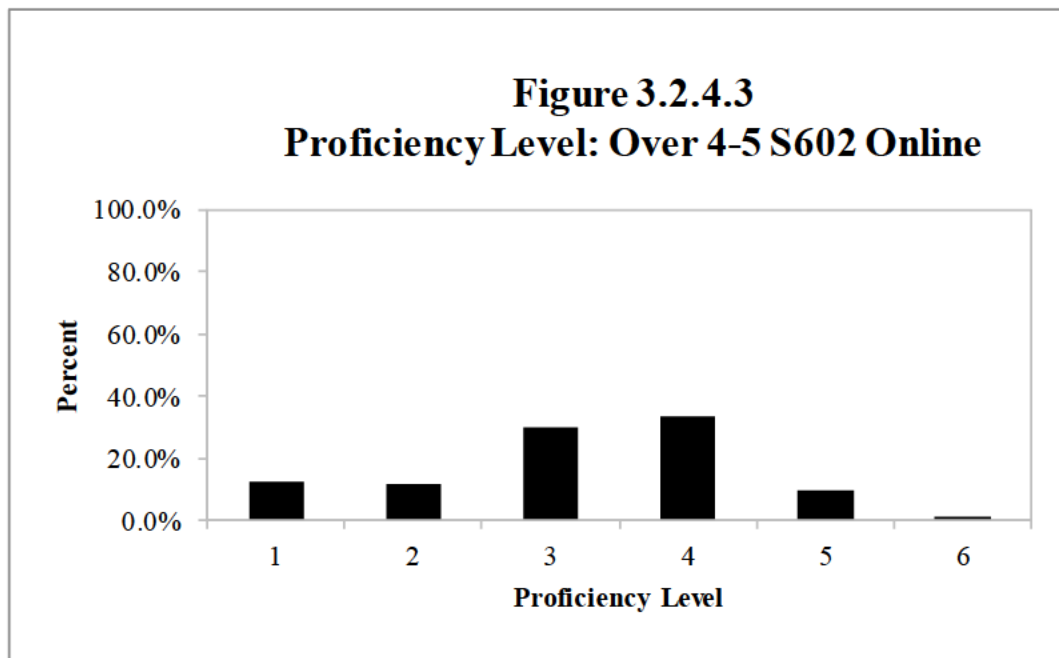
Table 3.2.4.3

Proficiency Level Distribution: Over 4–5 S602 Online

Level	G4 Count	G4 Percent	G5 Count	G5 Percent	Total Count	Total Percent
1	18,836	11.44%	19,608	14.03%	38,444	12.63%
2	20,117	12.22%	16,950	12.13%	37,067	12.18%
3	50,849	30.88%	41,645	29.80%	92,494	30.38%
4	55,277	33.57%	47,590	34.06%	102,867	33.79%
5	17,164	10.42%	12,636	9.04%	29,800	9.79%
6	2,429	1.48%	1,309	0.94%	3,738	1.23%
Total	164,672	100.00%	139,738	100.00%	304,410	100.00%

Figure 3.2.4.3

Proficiency Level: Over 4–5 S602 Online



3.2.4.4 Grades 6–8

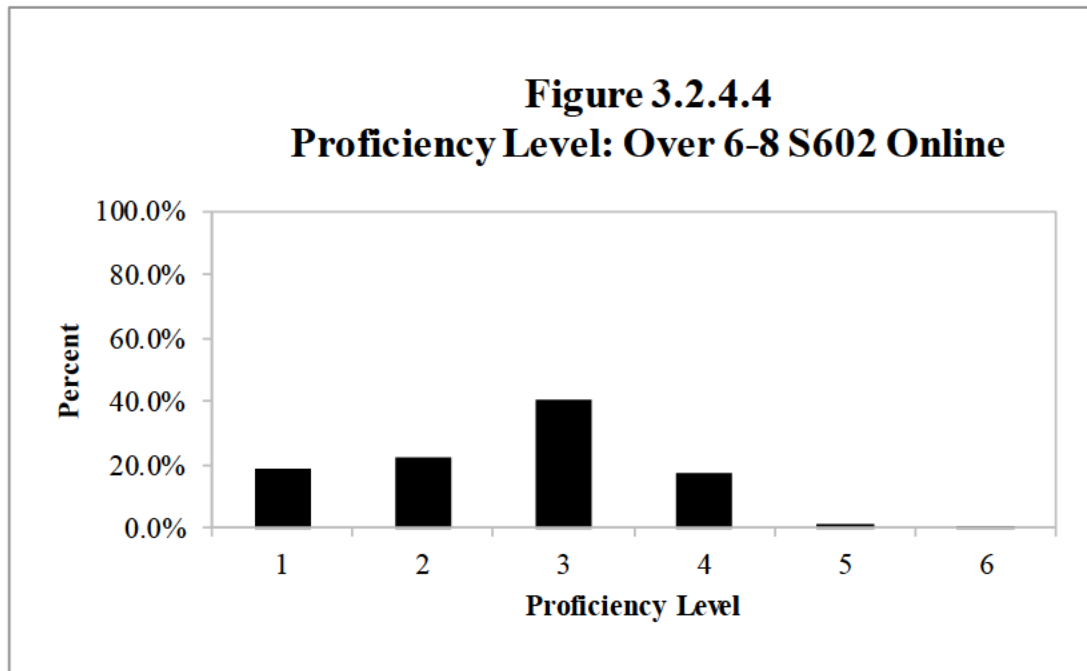
Table 3.2.4.4

Proficiency Level Distribution: Over 6–8 S602 Online

Level	G6 Count	G6 Percent	G7 Count	G7 Percent	G8 Count	G8 Percent	Total Count	Total Percent
1	19,087	16.51%	22,651	18.72%	24,979	20.52%	66,717	18.62%
2	27,759	24.01%	26,629	22.01%	24,730	20.32%	79,118	22.08%
3	51,495	44.54%	48,701	40.25%	45,620	37.48%	145,816	40.69%
4	16,475	14.25%	21,615	17.86%	24,562	20.18%	62,652	17.48%
5	758	0.66%	1,341	1.11%	1,772	1.46%	3,871	1.08%
6	49	0.04%	69	0.06%	54	0.04%	172	0.05%
Total	115,623	100.00%	121,006	100.00%	121,717	100.00%	358,346	100.00%

Figure 3.2.4.4

Proficiency Level: Over 6–8 S602 Online



3.2.4.5 Grades 9–12

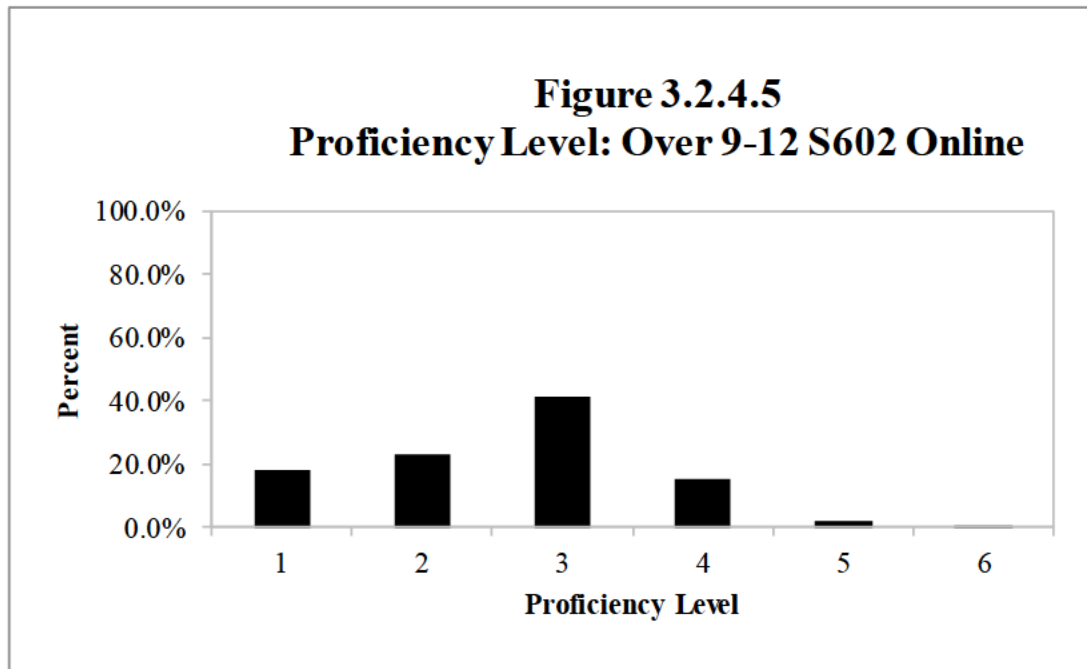
Table 3.2.4.5

Proficiency Level Distribution: Over 9–12 S602 Online

Level	G9 Count	G9 Percent	G10 Count	G10 Percent	G11 Count	G11 Percent	G12 Count	G12 Percent	Total Count	Total Percent
1	23,681	17.93%	19,502	17.07%	17,199	18.87%	13,540	20.00%	73,922	18.25%
2	29,044	21.99%	25,903	22.68%	21,523	23.62%	18,322	27.06%	94,792	23.40%
3	55,599	42.09%	47,990	42.01%	37,031	40.63%	27,114	40.05%	167,734	41.40%
4	21,224	16.07%	18,783	16.44%	13,777	15.12%	7,924	11.70%	61,708	15.23%
5	2,397	1.81%	1,985	1.74%	1,567	1.72%	799	1.18%	6,748	1.67%
6	135	0.10%	72	0.06%	38	0.04%	7	0.01%	252	0.06%
Total	132,080	100.00%	114,235	100.00%	91,135	100.00%	67,706	100.00%	405,156	100.00%

Figure 3.2.4.5

Proficiency Level: Over 9–12 S602 Online



4. Annual Updates of Validity Evidence

This section presents studies conducted as validity evidence for the WIDA ACCESS assessments. According to the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), validity is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed use. Particular interpretations for specified uses begin by specifying the construct the test is intended to measure. Rather than referring to distinct types of validity, the Standards refer to types of validity evidence. According to the Standards, the evidence can be based on (1) test content, (2) response processes, (3) internal structure, and (4) relation to other variables, which are listed in Section 4.1.

The validity evidence of the Standards is also observed in “A State’s Guide to the U.S. Department of Education’s Assessment Peer Review Process” document (Department of Education, 2018) to support states’ use of ELP assessments for reviewing of validity evidence, as well as being linked to the Assessment Use Argument (AUA) (Bachman & Palmer, 2010) to support the claims of validity of ACCESS Online assessment. WIDA structures its validity arguments using AUA model in lieu of the model highlighted in the *Standards for Educational and Psychological Testing*. The AUA has similar topics; however, they are organized differently. The following list contains a short summary of each AUA claim. For the full AUA validity claims please refer to the WIDA Assessment Use Argument document.

Claim 1 (Consequences): With the use of ACCESS, the intended decisions will have beneficial consequences for stakeholders, in terms of using ACCESS and the decisions made based on ACCESS.

Claim 2 (Decisions): Decisions based on ACCESS test results are made by individuals in a timely manner and affect a variety of stakeholders. Two types of decisions made based on ACCESS results are classification and programming decisions. The decisions take into consideration educational and societal values, and relevant laws, rules, and regulations, and they are equitable for the intended stakeholders.

Claim 3 (Interpretations): The interpretations of students’ academic English language proficiency in four domains are *relevant* to the classification, placement and programming decisions; *sufficient*, in conjunction with additional information as outlined in state and local policies, to make such decisions; *meaningful* with respect to the WIDA English Language Development (ELD) Standards; *generalizable* to the academic English language used in K–12 instructional settings, and *impartial* to all students.

Claim 4 (Assessment records: Scores): ACCESS scores are consistent across different aspects of test administration, different test tasks, and different groups of students. Test forms and metrics accurately represent the construct being measured and result in expected test taker performances.

4.1 Standards

4.1.1 Test Content

The relationship between the content of a test and the construct to be measured is called content validity. Test content includes the themes, wording, and format of the items, tasks, or questions on a test. Administration and scoring may also be part of the content. Empirical or logical evidence can show how appropriately the content reflects the domain as we interpret test scores.

4.1.2 Response Processes

Empirical analysis of how test takers process tests provide evidence of the nature between performance and the construct. Examples of this validity include analyzing individual item responses, different response processes in answering questions by subgroups or evaluating test-takers performance.

4.1.3 Internal Structure

Validity related to internal structure indicates how test items/components agree with the construct on which the score interpretation is based. The internal structure of the construct can be unidimensional or contain multidimensional components.

4.1.4 Relation to Other Variables

The interpretation of the test scores with an external indicator provides valuable validity evidence. We often ask how accurately the test score predicts the criterion variable. The test criterion validity has two different validities: concurrent and predictive validity. Predictive validity is how accurately test scores predict the future performance of criterion scores. Concurrent validity indicates how test scores relate to criterion scores at the same time.

4.2 Annual Validity Studies

4.2.1 Validating a New Writing Scoring Scale Using Multi-Faceted Rasch Analysis

Chuang, P-L. (2024, April). *Validating a new writing scoring scale using multi-faceted Rasch analyses* [Technical report]. WIDA, Wisconsin Center for Education Research, and the Board of Regents of the University of Wisconsin System. <https://wida.wisc.edu/resources/validating-new-writing-scoring-scale-using-multi-faceted-rasch-analyses>

This study is situated within a larger ongoing project of developing a new scoring scale for the ACCESS for ELLs Writing test. The project aims to validate the newly developed writing scoring scale by examining its feasibility in differentiating student ability and practical scoring use.

The development of rating scales can be mainly categorized into two approaches (Turner & Upshur, 2002). The first approach is theory-based and uses “theoretical views about the development of L2 ability” to develop scale descriptors. While these scales have strong theoretical support, they are often criticized as being irrelevant to the test task or unclear due to the use of relative wording. To address these issues, empirically-based scales are developed.

Once a rating scale has been developed, validation should be performed to ensure its quality and functionality. Scales can be validated quantitatively and qualitatively. A multi-faceted Rasch analysis is commonly performed to examine the psychometric properties of a rating scale. It combines different facets such as examinees, raters, scoring criteria, or test items into one analysis and converts raw scores into a logit interval scale (Linacre, 2004).

The study examines how the newly developed writing scoring scale functions by testing the following four hypotheses:

1. A well-functioning rating scale will result in all score points being used and no single score point being overly used (variation in ratings).
2. A well-functioning rating scale will result in small differences between raters in terms of their leniency and harshness as a group (rater separation).
3. A well-functioning rating scale will result in high rater reliability as indicated by rater point biserial correlations and exact agreement rates (rater reliability).
4. A well-functioning rating scale will result in high candidate discrimination (student discrimination).

This study shows the quality and benefits of empirically developing a writing scoring scale. The validation results suggest the scale’s ability to represent test takers of various proficiency levels and its capacity to help raters perform similarly to each other, likely because it captures a range of possible performances based on empirical data. Scale developers can consider adopting this approach to develop task-relevant scales to ensure more accurate scoring. This study also demonstrates the importance of multi-faceted Rasch analysis in validating a scoring scale for an operational writing test. The analysis provided meaningful information including rater severity and student discrimination, allowing for a comprehensive diagnosis of scale functionality. This method is not only applicable to large-scale assessments like ACCESS for ELLs but is also appropriate for smaller-scale local tests or classroom assessments for which sufficient data is collected.

4.2.2 Development of a New WIDA Writing Scoring Rubric for Grades 1-12

Chapman, M., Chuang, P., Bitterman, T., & Elliott, H. (2024, August). *Development of a new WIDA writing scoring rubric for grades 1-12* [Technical Report]. WIDA, Wisconsin Center for Education Research, and the Board of Regents of the University of Wisconsin System.

<https://wida.wisc.edu/sites/default/files/resource/Technical-Report-Development-New-WIDA-Writing-Scoring-Rubric-Grades-1-12.pdf>

The main aim of this project was to develop a new scoring rubric grounded in the WIDA English Language Development Standards Framework, 2020 Edition: Kindergarten–Grade 12 (hereafter, WIDA ELD Standards Framework, 2020 Edition or 2020 Edition). This rubric will be used for scoring responses to the writing tasks on ACCESS for ELLs Online, ACCESS for ELLs Paper, WIDA Screener Online, and WIDA Screener Paper.

Two features of the WIDA ELD Standards Framework, 2020 Edition, that differed from previous editions prompted the need for a new writing rubric. The first was the shift to grade-level cluster-specific proficiency level descriptors. The second was the greater emphasis on the discourse dimension of language in the 2020 Edition. The WIDA ELD Standards Framework has consistently described three dimensions of language: discourse, sentence, and word/phrase. In the 2020 Edition, the discourse dimension was expanded into three different criteria: organization of language, cohesion of language, and density of language.

The writing scoring rubric underwent multiple rounds of review and revisions via the processes described in the previous sections. Some of the major decisions made based on the input from these reviews were:

- The new writing scoring rubric features eight score points (0–7). A majority of reviewers offered support for the 0–7 raw score range, though some reviewers reported that score points 6 and 7 were difficult to distinguish and should be consolidated. Descriptors for these score points were revised to make them more distinguishable. For example, greater emphasis was placed on describing the extent to which responses demonstrated features of the intended key language uses (KLUs) and relevant content area.
- The plus score points (e.g., 4+) that were a feature of the WIDA Writing Scoring Scale Grades 1–12 are not included in the new WIDA Writing Scoring Rubric Grades 1–12. Reviewers, including internal WIDA reviewers, educators, and DRC reviewers, unanimously supported the removal of the plus score points in the new rubric. Reviewers commented that the shift away from using “+” in the score points would help make scoring more straightforward and may contribute to increased rater reliability.
- Score points 3 through 7 include three descriptors, one for each dimension of language encoded in the WIDA Standards. Score points 1 and 2 include one and two descriptors respectively, reflecting the observation that student responses at these score points tend largely to feature writing at the word/phrase (SP1) and sentence (SP2) dimensions. Discourse descriptors are typically not relevant to these responses.
- Educators requested that the new writing scoring rubric add more detail to the scoring notes and glossary sections. Guidance is now included on how to rate responses that include languages other than English in the rubric scoring notes for the first time.
- Reviewers consistently commented that the new scoring rubric is an improvement on the writing scoring scale, which will be easier to use operationally for both DRC raters and educators.

4.2.3 Examining English Learner Testing, Proficiency, and Growth: Before, During, and “After” the COVID-19 Pandemic

Sahakyan, N., & Poole, G. (2023, April). *Examining English learner testing, proficiency, and growth: Before, during, and “after” the COVID-19 pandemic* [Research report]. WIDA, Wisconsin Center for Educational Research, and the Board of Regents of the University of Wisconsin System. <https://wida.wisc.edu/resources/examining-english-learner-testing-proficiency-and-growth-before-during-and-after-covid-19>

This study shows how academic English proficiency has continued to decline, on average, for the overall population of English learners (ELs) since the COVID-19 pandemic. The analysis draws on aggregated individual-level data from the ACCESS for ELLs Online assessment, taken annually by students identified as ELs across the WIDA Consortium. We present the number of tested ELs, as well as their average proficiency and gain scores for the six academic years from 2017–2018 through 2022–2023. The 2024 report examines aggregate trends in English learner proficiency and growth since the pandemic, adding the most recent ACCESS assessment data from the 2022–2023 school year. It is also the first report in the series to disaggregate and present outcomes by English learner subgroup, drawing attention to persistent and growing disparities in the average proficiency of ELs identified as Hispanic and non-Hispanic.

Overall, our findings suggest that ELs in higher grade levels especially are showing slower growth than pre-pandemic averages. If students do not receive the supports they need to reach reclassification-level proficiency, many more are likely to receive the “long-term” label, which—in addition to further stigmatizing students identified as ELs—has implications for school and district accountability. Delayed language proficiency or reclassification may also contribute to additional barriers that many English learners face in accessing advanced coursework and academic milestones important for college and career readiness.

In addition to the overall trends in declining proficiency, disaggregated analyses by subgroup suggest that pandemic-related disruptions may have exacerbated some of the existing disparities within the English learner population, in particular between Hispanic and non-Hispanic English learners’ average outcomes. As many ELs continue to face disproportionately low rates of English language development, these analyses point to uneven barriers in their academic experiences—even after schools returned to in-person instruction. More nuanced analyses are needed to unpack and understand how different subgroups of students may have faced disproportionate challenges during the COVID-19 pandemic and how those challenges may continue to affect students in ongoing ways.

With the 2023–24 administration of ACCESS wrapped up across the consortium, WIDA research reports will continue to inform the national conversation around post-pandemic recovery and English learner outcomes in K–12 education. We recommend that states and districts conduct their own local analyses of overall and disaggregated student outcomes to determine what resources and supports are most appropriate to meet the unique needs of their students. In particular, administrators and policymakers might consider the potentially ongoing ways in which the pandemic may have exacerbated disparities within their community—

not only between emergent multilingual students and their peers, but also within the English learner population as well.

The most recent results can be found at

<https://wida.wisc.edu/sites/default/files/resource/Research-Report-Examining-English-Learner-Testing-Proficiency-Growth-2024.pdf>

5. Reliability

Following the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014), when interpreting test scores, it is important to evaluate their reliability, as the interpretation of test scores depends on the assumption that students exhibit some degree of consistency in their scores across independent administrations of the same testing procedure. We expect that students mastering the domain will consistently perform well, and those who have not mastered the domain will consistently perform less well, regardless of the sample of items and tasks used to assess students. Furthermore, because we assume that all items and tasks on such a test measure some aspect of the domain of interest, we expect that students will perform consistently across different items and tasks measuring the same ability within the test. Therefore, it is important to evaluate the degree to which students' test scores are consistent across replications of the same testing condition.

However, different samples of performances from the same student are rarely identical. A student's responses to sets of test items or tasks vary from one sample of test items or tasks targeting the domain to another, and from one occasion to another, even under strictly controlled conditions. In addition, different raters may award different scores to the same student's performance on a test task. These sources of variation are reflected in the students' scores. Therefore, it is important to evaluate the extent to which differences in students' test scores reflect true differences in the knowledge, skills, or abilities being tested, rather than fluctuations due to chance.

The reliability of the test scores depends on how much the scores vary across replications of the testing procedure, and analyses of reliability depend on the types of variability likely to be of concern in the testing procedure. There are several ways to collect reliability data and to estimate reliability, some of which depend on the exact nature of the measurement, the intended use of the test scores, the assessment design, and the potential sources of measurement error that might contribute to inconsistency in students' scores across different test administrations.

The reliability information presented in this section is organized to comply with Critical Element 4.1 of the Every Student Succeeds Act Peer Review requirements (U.S. Department of Education, 2018) and follows the guidelines of the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014). We present information regarding the reliability of the domain scale scores first, followed by information about the reliability of the composite scale scores.

Policymakers in states and districts use ACCESS Listening, Reading, Writing, and Speaking tests to determine the English language proficiency of students based on their scores in each of the four domains. Therefore, the main concern in interpreting these scores is how consistent the scores would be over replications of the same testing procedure. We use **internal consistency reliability statistics** to address this question (Section 5.1).

Additionally, for the Writing and Speaking domains, because having different raters evaluate the same students' responses to tasks may result in inconsistent scoring, a potential source of variation in those scores is the rater. In Section 5.2, we report the **interrater agreement** rates

that the raters achieved when evaluating students' responses to the Writing and Speaking tasks. We can use these statistics to determine how consistent the students' scores would have been if different raters had evaluated their responses. Since we use an item response theory (IRT)-based method to estimate students' **latent scores** (i.e., test scores based on variables that we cannot see or directly measure but which we can infer mathematically through advanced statistical techniques by using students' scores on variables that we can observe), we also examine the amount of **measurement error** in students' scores using the **conditional standard error of measurement** (CSEM) (Section 5.3). Lastly, in Section 5.4, we evaluate the reliability of the classifications of students into WIDA proficiency levels based on their domain scores (the most important interpretation of the test scores) in terms of the **accuracy and consistency** of the classification decisions made. In each subsection, we present detailed descriptions of the methods, data sources, and procedures.

Policymakers in states and districts use ACCESS **composite scale scores** to describe the English language proficiency of students in the respective composites. Therefore, the most important concern in interpreting these scores is how consistent the scores would be over replications of the same testing procedure. We use internal consistency reliability statistics to address this question, and in Section 5.5 we provide the results. In addition, in Section 5.6, we examine the CSEM of these scores. Lastly, in Section 5.7, we evaluate the reliability of the classifications in terms of the accuracy and consistency of the decisions made about students' levels of English language proficiency based on their composite scale scores. In each subsection, we present detailed descriptions of the methods, data sources, and procedures.

Internal Consistency Reliability Statistics: One way to evaluate the consistency of students' test scores across test administrations is to examine how the students would have performed on alternate forms of the same test (i.e., **parallel test form reliability**). Given our assumption that the ability the test measures is constant for each student over two administrations of alternate forms, the more variation found across the two administrations, the more evidence for lower reliability. The **measurement error** represents the sources of inconsistency across the two administrations, taken together. We consider measurement error to be random and to occur by chance. For example, there may be some construct-irrelevant knowledge and/or skills that some items or tasks measure that affect students' scores but are not part of the ability that the test intends to measure.

Unless students take two alternate versions of the same test, we cannot calculate test score reliability directly. Thus, we usually estimate it from student responses to a single form of the test. Methods employed to estimate reliability using test scores from a single test administration are based on classical test theory and are referred to as estimates of **internal consistency**. An internal consistency reliability statistic is a useful estimate of alternate-forms reliability, providing an estimate of the consistency of students' performances across items and tasks within a test. The most common index of internal consistency reliability is **Cronbach's coefficient alpha** (Cronbach, 1951), which is a lower-bound estimate of test reliability. Conceptually, we think of Cronbach's coefficient alpha as the correlation obtained between performances on two halves of the same test if every possible way of dividing the test items and tasks in two was attempted. Because Cronbach's coefficient alpha is a correlation of students' performances on all possible pairs of test items and tasks, it may be low if some items

or tasks are measuring something other than what most of the other items and tasks are measuring (and thus leading to inconsistent student performances). In this way, Cronbach's coefficient alpha expresses how well the items and tasks on a test appear to measure the same ability. The Cronbach's coefficient alpha of internal consistency ranges from 0 to 1. If students achieve their scores by a completely random process (i.e., their scores are not correlated or share no covariance), then the reliability estimate is very close to 0. On the other hand, if students' scores are perfectly consistent (i.e., their scores have high covariances), then the internal consistency coefficient will approach 1.

While there is no one set of criteria that the testing community uses when interpreting Cronbach's coefficient alpha values, from time to time, researchers have proposed various arbitrary criteria that one could apply. Initially, Cronbach (1951) argued that it was 'desirable' to have a high alpha value for an instrument that test developers were using to report individual scores since the scores on that instrument needed to be interpretable, and that would require a high alpha value. Later, Nunnally (1978) suggested that researchers should consider a value of 0.70 as an acceptable lower limit if they were engaged in the early stages of research (e.g. when developing a scale). Today, it has become common practice to cite Nunnally's suggested 0.70 criterion as a minimum acceptable lower limit for this value for all types of research. However, in so doing, researchers ignore Nunnally's more nuanced guidance: If researchers were engaged in basic research, Nunnally advised that they should use a higher cut-off value (i.e., 0.80 or higher), and those engaged in applied research should use a much higher cut-off value (0.90 or higher) (Lance et al., 2006). Since Nunnally's time, some researchers have suggested even more nuanced interpretations of various alpha values. For example, George and Mallery (2003) proposed the following interpretations: " ≥ 0.90 – Excellent, ≥ 0.80 – Good, ≥ 0.70 – Acceptable, ≥ 0.60 – Questionable, ≥ 0.50 – Poor, and ≤ 0.50 – Unacceptable" (p. 231). There is little consensus among the experts in their views of what the acceptable lower limit of Cronbach's coefficient alpha value should be, or for that matter, how one should interpret various values. This lack of consensus led the authors of the *Standards for Educational and Psychological Measurement* (2014) to conclude, "The choice of [reliability/precision] estimation and the minimum acceptable level for any index remain a matter of professional judgment" (p. 41). For this report then, WIDA has made the decision that within the domains of Listening, Reading, and Speaking, an alpha value of ≥ 0.80 is acceptable, while an alpha value of ≥ 0.65 is acceptable for the Writing domain.

Reliability statistics such as the Cronbach's coefficient alpha of internal consistency are affected by two factors: (1) the number of test items or tasks, and (2) the total number of score points students achieve. That is, all things being equal, the greater the number of items or tasks measuring the same ability there are on the test, the higher the internal consistency reliability statistics. Additionally, because reliability statistics refer to the consistency of scores *for a group of students*, the distribution of that specific group's ability measures affects these statistics. If the students in the group are nearly equal in the ability that the test measures (i.e., their scores are concentrated in the center of the ability distribution), small changes in their scores can easily change their relative positions in the group. Consequently, the internal consistency reliability statistics will be low. In this case, the statistics may be telling us more about the group of students tested than about the test itself. On the other hand, if the students

in the group differ widely in the ability that the test measures (i.e., their scores are distributed across the ability continuum), small changes in their scores will not affect their relative positions in the group as much, and the internal consistency reliability statistics will be higher. Therefore, reliability can be as much a function of the performance of test items and tasks as of the performance of the sample of students tested. That is, the same test can produce widely disparate reliability indices based on the ability distribution of the group of students. This means, in turn, that when interpreting estimates of internal consistency, it is wise to keep in mind the specific set of test items and tasks and the distribution of ability measures in the group of students used in the estimation.

Interrater Agreement: The raters' behavior is a potential source of variance in students' scores for the productive domains of ACCESS (i.e., Writing and Speaking). ACCESS scoring procedures, rater training, and quality control monitoring processes are described elsewhere in this report (see Part 1, Section 4). In Section 5.2, we report the **interrater agreement rates** for scoring students' responses to the Writing and Speaking tasks. These values reflect how consistent the students' scores would be if different groups of raters scored their responses. Additionally, in this section of the report, we present a detailed description of the methods, data sources, and procedures we used when calculating interrater agreement rates.

Measurement Error: In addition to evaluating test score reliability in terms of estimates of internal consistency, we can calculate the amount of measurement error in students' test scores in two different ways. One way is to hypothesize that there is an error-free measure of each student's true ability, referred to as the **true score** in classical test theory. The true score is a theoretical value, so it is not a known quantity. Rather, we view it as the hypothetical average score over repeated replications of the same testing condition (Livingston, 2018, p. 9). Under the assumptions of classical test theory, the **error of measurement** over a replication of a testing condition provides an estimate of the amount of variability from students' true scores that we would expect. In practical testing contexts, it is generally not possible to replicate a testing condition (i.e., have students take the same test form multiple times), so it is not possible to estimate the standard error of each student's score using a repeated measures design. Instead, we calculate the average error of measurement over the population of students who take the test, and then we use that as an indication of the amount of variation in any individual student's score that we would expect. Classical test theory refers to this average as the **standard error of measurement** (SEM), which indicates how much students' scores differ from their true scores, on average, on the raw score metric. Because it is a standard deviation of the distribution of errors of measurement, we can construct a **confidence interval** to indicate how the errors of measurement are affecting the scores. Test scores with large SEMs pose a challenge to the interpretation of the reliability of any single test score.

A second way to address the impact of measurement errors on students' test scores is to estimate the SEM for specific scores using IRT. IRT addresses reliability using the **test information function**, which indicates the precision with which we can use student performances on items and tasks to estimate the **latent** (i.e., true) **ability** of each student (i.e., **latent scores**). The square root of the inverse of the information function at any point on the latent ability distribution is the **conditional standard error of measurement** (CSEM). The CSEM provides information about the amount of error we would expect in any student's score

at that point on the underlying latent ability scale, which IRT refers to in terms of the **latent score metric** (i.e., the IRT metric for expressing student ability, as opposed to the raw score metric). In addition, by using IRT, we can estimate indices analogous to traditional reliability coefficients such as Cronbach's coefficient alpha from the test information function and the distribution of the latent scores in the same student population.

Classification Accuracy and Consistency: One of the main purposes of the WIDA ACCESS program is to identify the English language proficiency levels of students concerning the WIDA ELD Standards. Because of the emphasis on the classification of student performance into six WIDA proficiency levels, it is important to know how consistently ACCESS scores do indeed classify students into those proficiency levels (American Educational Research Association et al., 2014). The questions that we want to answer are different from the questions that the reliability coefficient answers. Instead of looking at the reliability of a specific student score, we want to know the consistency of the decisions we make when we use students' test scores to classify them into a smaller number of proficiency levels. One way to approach this question is to estimate the degree to which the classification decisions we are making based on the students' **observed test scores** agree with the classification decisions we would make based on students' **theoretical true scores**. This estimate is known as **decision accuracy**. A second way to approach this question is to estimate the degree to which the classification decisions we are making based on the students' test scores agree with the classification decisions we would make based on students' scores on an alternate form of the test. This estimate is known as **decision consistency**.

5.1 Reliabilities of the Domain Scores

Listening and Reading: Internal consistency statistics based on classical test theory are applicable only for a fixed-length test where all students take the same set of test items (Thissen, 2000). For the Listening and Reading tests, which are computer adaptive, we cannot compute traditional internal consistency reliabilities because not all students take the same set of items. We estimate the reliabilities of students' domain scale scores for Listening and Reading by grade-level cluster using an IRT-based **marginal reliability method** that Thissen (2000) derived. Unlike the traditional internal consistency statistics that are based on students' raw scores, the marginal reliability method for calculating reliability uses students' domain scale scores and the distribution of the students' domain scale scores on the theta scale (i.e., **domain theta scores**) in its estimation. However, we can interpret the marginal reliability coefficient like other traditional internal consistency coefficients such as Cronbach's coefficient alpha (Thissen, 2000).

The formula for calculating an IRT-based marginal reliability coefficient using the method that Thissen (2000) developed is

$$\bar{\rho} = \frac{\sigma_{\theta}^2 - \text{average}(CSEM_{\text{observed}}^2)}{\sigma_{\theta}^2}$$

where

$\bar{\rho}$ is the average reliability

σ_{θ}^2 is the variance of the distribution of the students' domain theta scores

$CSEM_{observed}^2$ is the squared observed CSEM for each student's domain theta score.

We can calculate the IRT-based marginal reliability coefficient directly (Thissen, 2000); however, it is computationally intensive. Since this estimate is equivalent to the **Rasch student separation reliability coefficient** (Linacre, 1999), which is regularly reported as part of the output from a Winsteps analysis, for purposes of efficiency WIDA chose to report the Rasch student separation reliability coefficients as the test score reliability estimates for the Listening and Reading domains. The Rasch student separation reliability coefficient is an estimate of the ratio of "true measure variance" to "observed measure variance" (Linacre, 1999). The student separation reliability coefficient answers these questions: How consistent are the students' relative positions in the group tested, as indicated by their domain scale scores? How reproducible is the student ability measure order of this sample of students for this set of items? The more the students differ in ability, the less likely that small changes in their domain scale scores will affect their relative positions in the group, and the higher the student separation reliability coefficient will be. Thus, to obtain high student separation reliability, a wide sample of student ability in the domain (i.e., a large student ability range) and/or low measurement error (i.e., a test containing many items) is required (Linacre, 2020). Student separation reliabilities can range from 0.00 to 1.00. A student separation reliability < 0.80 implies that the test may not be sensitive enough to distinguish between high- and low-performing students, and thus more items may be needed (Linacre, 2020). To obtain these values, we used the item parameters and population student data as inputs for the Winsteps program.

The tables in Section 5.1.1 present test score reliability information for the Listening domain, while the tables in Section 5.1.2 present test score reliability information for the Reading domain. For these two domains, we provide the Rasch student separation reliability coefficients that are based on students' ACCESS Online domain theta scores. For each of these domains, we present four tables. The first table reports the Rasch student separation reliability coefficient (labeled as 'Rasch Student Separation Reliability Coefficient' in the table) for all students in each grade-level cluster. Each row in the table represents a grade-level cluster, and values for the numbers of students, numbers of items, and the student separation reliability estimate are provided based on students' domain theta scores in each grade-level cluster. The second table provides the same information for the population of female students and the population of male students. The third table provides information by ethnicity, for Hispanic and non-Hispanic students, and the fourth table provides information for the population of students who have an individualized education plan (IEP).

For Listening, the Rasch student separation reliability coefficients based on the domain theta scores for all students ranged from 0.86 to 0.89 across the grade-level clusters (Table 5.1.1.1). The Rasch student separation reliability coefficients ranged from 0.86 to 0.89 for male students; 0.86 to 0.89 for female students (Table 5.1.1.2); 0.86 to 0.89 for Hispanic students; 0.84 to 0.88 for non-Hispanic students (Table 5.1.1.3). For students with an IEP, the Rasch student separation reliability coefficients ranged from 0.79 to 0.89 for students with an IEP (Table 5.1.1.4).

For Reading, as shown in Table 5.1.2.1, the Rasch student separation reliability coefficients based on the domain theta scores for all students ranged from 0.85 to 0.90 across the grade-level clusters. The Rasch student separation reliability coefficients ranged from 0.85 to 0.90 for male students; 0.86 to 0.90 for female students (Table 5.1.2.2); 0.82 to 0.89 for Hispanic students; 0.88 to 0.91 for non-Hispanic students (Table 5.1.2.3); and 0.80 to 0.87 for students with an IEP (Table 5.1.2.4).

Writing and Speaking: Cronbach's coefficient alpha is widely used as an estimate of reliability, particularly for the internal consistency of test items and/or tasks, and this statistic is appropriate for calculating the reliabilities of students' scores from the administration of the fixed forms of the Writing and Speaking tests. Conceptually, we can think of it as the correlation obtained between students' performances on two halves of the Writing or Speaking test if every possible way of dividing the test tasks in two was attempted. Thus, Cronbach's coefficient alpha may be low if some tasks are measuring something other than what the majority of the tasks are measuring. In this way, Cronbach's coefficient alpha expresses how well the tasks on a test appear to measure the same ability.

The formula for calculating Cronbach's coefficient alpha for the fixed forms of the Writing and Speaking tests is

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_t^2} \right]$$

where

n = the number of tasks

σ_i^2 = the variance of students' raw scores on task i

σ_t^2 = the variance of students' total raw scores.

For the Writing and Speaking tests, tables in this section also present the SEM, a single value for estimating the errors of measurement in students' raw scores calculated using a classical test theory-based approach. It is a function of two statistics: (1) the Cronbach's coefficient alpha calculated using students' raw scores on the test, and (2) the (observed) standard deviation (SD) of the students' total raw scores. It is on the raw score metric. The Cronbach's coefficient alpha is calculated as

$$SEM = SD \sqrt{1 - reliability}$$

Since the SEM is an estimate of the standard deviation of the distribution of measurement errors, we can use the SEM to create a band around a student's observed raw score. Under the assumption that the error of measurement follows a normal distribution, the student's true score would lie with a certain degree of probability within this band. Statistically speaking, then, there is an expectation that a student's true raw score has a 68% probability of falling within the band extending from the observed score minus 2 SEMs to the observed score plus 2 SEMs. Since SEMs are expressed on the raw score metric, it is wise to keep the range of the possible

raw score distribution in mind when interpreting the SEM. For example, if the Online Writing test has a possible raw score range of 0 to 18 and one SEM equals 2 score points, and if a student receives a score of 10 on the test, we know with 95% certainty that the student's true score lies somewhere between a raw score of 8 and 12 (i.e., 10 minus, or plus, 2 SEMs). Similarly, if one SEM equals 1 score point, we would say with 68% certainty that the student's true score lies between 9 and 11 (i.e., 10 minus, or plus, 1 SEM). The smaller the value of the SEM, the more precise the test scores will be.

The range of total possible raw score points for the Writing forms is 0 to 18. The ranges of total possible raw score points for the Speaking forms are 0 to 6 for Tier Pre-A, 0 to 18 for Tier A, and 0 to 24 for Tier B/C.

The tables in Section 5.1.3 present reliability information for the Writing test, and the tables in Section 5.1.4 present reliability information for the Speaking test. For these two domains, the tables report the number of tasks, the Cronbach's coefficient alphas, and the SEMs for all students and subgroups as the Every Student Succeeds Act Peer Review requires, thus facilitating the comparison of the reliability estimates computed based on the performance of individual subgroups to those computed based on the performance of all students. For each of these domains, we present four tables. The first table provides the Cronbach's coefficient alphas and the SEMs for all students based on their raw scores. Each row in the table represents a specific grade-level cluster and test form. For each form, the tables provide the number of students, number of tasks, total possible raw score points, Cronbach's coefficient alpha, and SEM. The second table provides the same information for the population of female students and the population of male students. The third table provides information by ethnicity, for Hispanic and Other students, and the fourth table provides information for the population of students who have an IEP.

Note that students' Writing reported scores are based on their performances on only two tasks starting with Online Series 501. Therefore, the Cronbach's coefficient alpha for the Writing domain may be lower than when estimated based on student performances on three tasks, as in the earlier series.

Writing Tier A: The Writing Tier A Cronbach's coefficient alphas computed based on the raw scores for all students ranged from 0.87 to 0.90. The Writing Tier A Cronbach's coefficient alphas ranged from 0.87 to 0.90 for male students; 0.87 to 0.90 for female students; 0.87 to 0.90 for Hispanic students; 0.85 to 0.89 for Other students; and 0.82 to 0.88 for students with an IEP.

Writing Tier B/C: The Writing Tier B/C Cronbach's coefficient alphas computed based on the raw scores for all students ranged from 0.68 to 0.77. The Writing Tier B/C Cronbach's coefficient alphas ranged from 0.68 to 0.78 for male students; 0.68 to 0.76 for female students; 0.69 to 0.78 for Hispanic students; 0.65 to 0.74 for Other students; and 0.69 to 0.82 for students with an IEP.

Speaking Tier Pre-A: The Speaking Tier Pre-A Cronbach's coefficient alphas computed based on the raw scores for all students ranged from 0.86 to 0.88. The Cronbach's coefficient alphas ranged from 0.86 to 0.88 for male students; 0.86 to 0.88 for female students; 0.86 to

0.88 for Hispanic students; 0.86 to 0.90 for Other students; and 0.85 to 0.93 for students with an IEP.

Speaking Tier A: The Speaking Tier A Cronbach's coefficient alphas computed based on the raw scores for all students ranged from 0.86 to 0.88. The Cronbach's coefficient alphas ranged from 0.85 to 0.88 for male students; 0.86 to 0.89 for female students; 0.86 to 0.89 for Hispanic students; 0.82 to 0.86 for Other students; and 0.80 to 0.88 for students with an IEP.

Speaking Tier B/C: The Speaking Tier B/C Cronbach's coefficient alphas computed based on the raw scores for all students ranged from 0.84 to 0.89. The Cronbach's coefficient alphas ranged from 0.84 to 0.88 for male students; 0.84 to 0.88 for female students; 0.85 to 0.89 for Hispanic students; 0.84 to 0.86 for Other students; and 0.84 to 0.89 for students with an IEP.

5.1.1 Listening

Table 5.1.1.1

Reliabilities of Domain Scores: List S602 Online

Cluster	# of Students	# of Items	Rasch Student Separation Reliability Coefficient
1	213,555	54	0.89
2-3	436,928	54	0.88
4-5	373,317	54	0.87
6-8	435,520	54	0.87
9-12	493,338	54	0.86

Table 5.1.1.2

Reliabilities of Domain Scores: List S602 Online by Gender

Cluster	# of Items	Gender	# of Students	Rasch Student Separation Reliability Coefficient
1	54	F	87,412	0.89
1	54	M	93,807	0.89
2-3	54	F	179,545	0.87
2-3	54	M	194,115	0.88
4-5	54	F	146,397	0.87
4-5	54	M	168,273	0.88
6-8	54	F	166,815	0.87
6-8	54	M	199,923	0.87
9-12	54	F	186,974	0.86
9-12	54	M	228,534	0.86

Table 5.1.1.3**Reliabilities of Domain Scores: List S602 Online by Ethnicity**

Cluster	# of Items	Ethnicity	# of Students	Rasch Student Separation Reliability Coefficient
1	54	H	142,577	0.89
1	54	O	65,369	0.88
2-3	54	H	295,236	0.87
2-3	54	O	130,952	0.87
4-5	54	H	253,301	0.88
4-5	54	O	105,871	0.85
6-8	54	H	303,448	0.87
6-8	54	O	108,695	0.85
9-12	54	H	345,145	0.86
9-12	54	O	121,349	0.84

Table 5.1.1.4**Reliabilities of Domain Scores: List S602 Online by IEP Status**

Cluster	# of Students	# of Items	Rasch Student Separation Reliability Coefficient
1	18,649	54	0.89
2-3	42,657	54	0.86
4-5	44,318	54	0.85
6-8	56,978	54	0.83
9-12	58,335	54	0.79

5.1.2 Reading**Table 5.1.2.1****Reliabilities of Domain Scores: Read S602 Online**

Cluster	# of Students	# of Items	Rasch Student Separation Reliability Coefficient
1	223,101	72	0.85
2-3	452,156	72	0.88
4-5	374,121	72	0.89
6-8	446,485	72	0.89
9-12	489,225	72	0.90

Table 5.1.2.2**Reliabilities of Domain Scores: Read S602 Online by Gender**

Cluster	# of Items	Gender	# of Students	Rasch Student Separation Reliability Coefficient
1	72	F	90,316	0.86
1	72	M	98,622	0.85
2-3	72	F	183,989	0.88
2-3	72	M	202,353	0.88
4-5	72	F	145,735	0.89
4-5	72	M	170,041	0.90
6-8	72	F	169,194	0.88
6-8	72	M	206,977	0.89
9-12	72	F	183,706	0.90
9-12	72	M	228,597	0.90

Table 5.1.2.3**Reliabilities of Domain Scores: Read S602 Online by Ethnicity**

Cluster	# of Items	Ethnicity	# of Students	Rasch Student Separation Reliability Coefficient
1	72	H	149,372	0.82
1	72	O	67,833	0.88
2-3	72	H	305,677	0.87
2-3	72	O	135,227	0.89
4-5	72	H	254,076	0.89
4-5	72	O	105,641	0.90
6-8	72	H	311,579	0.88
6-8	72	O	110,481	0.89
9-12	72	H	343,190	0.89
9-12	72	O	118,846	0.91

Table 5.1.2.4**Reliabilities of Domain Scores: Read S602 Online by IEP Status**

Cluster	# of Students	# of Items	Rasch Student Separation Reliability Coefficient
1	19,971	72	0.80
2-3	44,629	72	0.84
4-5	45,056	72	0.86
6-8	59,437	72	0.84
9-12	58,222	72	0.87

5.1.3 Writing**Table 5.1.3.1****Reliabilities of Domain Scores: Writ S602 Online**

Cluster	Tier	# of Students	# of Tasks	Total Possible Raw Score Points	Cronbach's Alpha	SEM
1	A	209,593	2	0-18	0.87	1.10
1	B/C	26,141	2	0-18	0.68	1.32
2-3	A	161,160	2	0-18	0.90	1.10
2-3	B/C	323,662	2	0-18	0.77	1.19
4-5	A	109,214	2	0-18	0.89	1.06
4-5	B/C	272,473	2	0-18	0.75	1.16
6-8	A	206,997	2	0-18	0.89	1.04
6-8	B/C	262,136	2	0-18	0.71	1.07
9-12	A	192,004	2	0-18	0.89	1.09
9-12	B/C	319,936	2	0-18	0.70	1.25

Table 5.1.3.2**Reliabilities of Domain Scores: Writ S602 Online by Gender**

Cluster	Tier	# of Tasks	Total Possible Raw Score Points	Gender	# of Students	Cronbach's Alpha	SEM
1	A	2	0-18	F	84,252	0.87	1.12
1	A	2	0-18	M	92,741	0.87	1.10
1	B/C	2	0-18	F	11,257	0.68	1.31
1	B/C	2	0-18	M	11,212	0.68	1.34
2-3	A	2	0-18	F	62,458	0.90	1.11
2-3	A	2	0-18	M	73,768	0.90	1.10
2-3	B/C	2	0-18	F	134,959	0.76	1.18
2-3	B/C	2	0-18	M	141,888	0.78	1.20
4-5	A	2	0-18	F	40,359	0.89	1.08
4-5	A	2	0-18	M	51,415	0.89	1.05
4-5	B/C	2	0-18	F	107,999	0.72	1.15
4-5	B/C	2	0-18	M	121,352	0.76	1.18
6-8	A	2	0-18	F	76,095	0.89	1.04
6-8	A	2	0-18	M	97,527	0.89	1.04
6-8	B/C	2	0-18	F	101,480	0.69	1.06
6-8	B/C	2	0-18	M	119,218	0.72	1.08
9-12	A	2	0-18	F	69,351	0.89	1.10
9-12	A	2	0-18	M	92,473	0.90	1.09
9-12	B/C	2	0-18	F	123,193	0.68	1.24
9-12	B/C	2	0-18	M	146,522	0.71	1.26

Table 5.1.3.3**Reliabilities of Domain Scores: Writ S602 Online by Ethnicity**

Cluster	Tier	# of Tasks	Total Possible Raw Score Points	Ethnicity	# of Students	Cronbach's Alpha	SEM
1	A	2	0-18	H	146,037	0.87	1.10
1	A	2	0-18	O	57,959	0.86	1.11
1	B/C	2	0-18	H	11,347	0.69	1.33
1	B/C	2	0-18	O	14,150	0.65	1.32
2-3	A	2	0-18	H	120,424	0.90	1.10
2-3	A	2	0-18	O	35,344	0.89	1.11
2-3	B/C	2	0-18	H	206,473	0.78	1.21
2-3	B/C	2	0-18	O	110,475	0.73	1.16
4-5	A	2	0-18	H	78,721	0.89	1.06
4-5	A	2	0-18	O	23,136	0.87	1.10
4-5	B/C	2	0-18	H	180,520	0.75	1.16
4-5	B/C	2	0-18	O	84,613	0.74	1.17
6-8	A	2	0-18	H	149,538	0.89	1.04
6-8	A	2	0-18	O	42,043	0.85	1.07
6-8	B/C	2	0-18	H	177,471	0.71	1.06
6-8	B/C	2	0-18	O	74,475	0.71	1.08
9-12	A	2	0-18	H	140,796	0.89	1.09
9-12	A	2	0-18	O	36,349	0.88	1.11
9-12	B/C	2	0-18	H	217,504	0.70	1.24
9-12	B/C	2	0-18	O	88,842	0.68	1.27

Table 5.1.3.4**Reliabilities of Domain Scores: Writ S602 Online by IEP Status**

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha	SEM
1	A	20,013	2	0-18	0.87	1.07
1	B/C	1,067	2	0-18	0.78	1.31
2-3	A	23,196	2	0-18	0.88	1.13
2-3	B/C	24,661	2	0-18	0.82	1.28
4-5	A	19,344	2	0-18	0.86	1.12
4-5	B/C	26,328	2	0-18	0.79	1.20
6-8	A	35,264	2	0-18	0.82	1.07
6-8	B/C	26,919	2	0-18	0.74	1.10
9-12	A	23,838	2	0-18	0.86	1.09
9-12	B/C	36,907	2	0-18	0.69	1.26

5.1.4 Speaking**Table 5.1.4.1****Reliabilities of Domain Scores: Spek S602 Online**

Cluster	Tier	# of Students	# of Tasks	Total Possible Raw Score Points	Cronbach's Alpha	SEM
1	Pre-A	14,805	3	0-6	0.87	0.82
1	A	108,591	6	0-18	0.88	1.36
1	B/C	91,009	6	0-24	0.84	1.60
2-3	Pre-A	29,967	3	0-6	0.88	0.77
2-3	A	142,324	6	0-18	0.88	1.34
2-3	B/C	273,300	6	0-24	0.85	1.56
4-5	Pre-A	12,656	3	0-6	0.86	0.82
4-5	A	76,302	6	0-18	0.87	1.28
4-5	B/C	288,771	6	0-24	0.85	1.59
6-8	Pre-A	30,892	3	0-6	0.87	0.77
6-8	A	107,825	6	0-18	0.86	1.37
6-8	B/C	309,360	6	0-24	0.87	1.56
9-12	Pre-A	38,194	3	0-6	0.87	0.70
9-12	A	219,027	6	0-18	0.86	1.34
9-12	B/C	243,942	6	0-24	0.88	1.48

Table 5.1.4.2**Reliabilities of Domain Scores: Spek S602 Online by Gender**

Cluster	Tier	# of Tasks	Total Possible Raw Score Points	Gender	# of Students	Cronbach's Alpha	SEM
1	Pre-A	3	0-6	F	5,969	0.88	0.80
1	Pre-A	3	0-6	M	6,662	0.86	0.85
1	A	6	0-18	F	42,643	0.89	1.34
1	A	6	0-18	M	49,478	0.88	1.36
1	B/C	6	0-24	F	39,337	0.85	1.60
1	B/C	6	0-24	M	38,000	0.84	1.60
2-3	Pre-A	3	0-6	F	11,750	0.88	0.76
2-3	Pre-A	3	0-6	M	13,704	0.88	0.77
2-3	A	6	0-18	F	56,224	0.88	1.33
2-3	A	6	0-18	M	64,576	0.88	1.34
2-3	B/C	6	0-24	F	114,992	0.84	1.56
2-3	B/C	6	0-24	M	119,675	0.85	1.56
4-5	Pre-A	3	0-6	F	4,892	0.87	0.81
4-5	Pre-A	3	0-6	M	5,856	0.86	0.83
4-5	A	6	0-18	F	28,524	0.87	1.28
4-5	A	6	0-18	M	35,766	0.87	1.28
4-5	B/C	6	0-24	F	114,462	0.85	1.60
4-5	B/C	6	0-24	M	128,661	0.85	1.60
6-8	Pre-A	3	0-6	F	12,074	0.87	0.75
6-8	Pre-A	3	0-6	M	13,756	0.86	0.78
6-8	A	6	0-18	F	39,482	0.86	1.37
6-8	A	6	0-18	M	51,006	0.85	1.36
6-8	B/C	6	0-24	F	118,435	0.87	1.57
6-8	B/C	6	0-24	M	142,543	0.86	1.56
9-12	Pre-A	3	0-6	F	13,985	0.86	0.69
9-12	Pre-A	3	0-6	M	18,231	0.87	0.70
9-12	A	6	0-18	F	81,486	0.86	1.35
9-12	A	6	0-18	M	103,461	0.87	1.33
9-12	B/C	6	0-24	F	93,392	0.88	1.50
9-12	B/C	6	0-24	M	112,126	0.89	1.46

Table 5.1.4.3**Reliabilities of Domain Scores: Spek S602 Online by Ethnicity**

Cluster	Tier	# of Tasks	Total Possible Raw Score Points	Ethnicity	# of Students	Cronbach's Alpha	SEM
1	Pre-A	3	0-6	H	11,269	0.87	0.83
1	Pre-A	3	0-6	O	2,884	0.86	0.76
1	A	6	0-18	H	78,369	0.89	1.35
1	A	6	0-18	O	27,321	0.86	1.37
1	B/C	6	0-24	H	54,038	0.85	1.59
1	B/C	6	0-24	O	34,972	0.84	1.61
2-3	Pre-A	3	0-6	H	22,574	0.88	0.78
2-3	Pre-A	3	0-6	O	5,893	0.87	0.69
2-3	A	6	0-18	H	105,922	0.88	1.34
2-3	A	6	0-18	O	32,478	0.84	1.34
2-3	B/C	6	0-24	H	172,805	0.85	1.55
2-3	B/C	6	0-24	O	94,883	0.84	1.57
4-5	Pre-A	3	0-6	H	9,088	0.86	0.83
4-5	Pre-A	3	0-6	O	1,892	0.88	0.70
4-5	A	6	0-18	H	54,940	0.87	1.29
4-5	A	6	0-18	O	16,442	0.82	1.28
4-5	B/C	6	0-24	H	192,406	0.85	1.58
4-5	B/C	6	0-24	O	88,599	0.84	1.61
6-8	Pre-A	3	0-6	H	22,443	0.86	0.78
6-8	Pre-A	3	0-6	O	4,336	0.88	0.65
6-8	A	6	0-18	H	77,767	0.86	1.37
6-8	A	6	0-18	O	21,976	0.82	1.37
6-8	B/C	6	0-24	H	212,454	0.87	1.56
6-8	B/C	6	0-24	O	84,564	0.86	1.58
9-12	Pre-A	3	0-6	H	29,163	0.86	0.71
9-12	Pre-A	3	0-6	O	5,190	0.90	0.58
9-12	A	6	0-18	H	158,107	0.87	1.34
9-12	A	6	0-18	O	46,975	0.83	1.32
9-12	B/C	6	0-24	H	163,839	0.89	1.48
9-12	B/C	6	0-24	O	70,039	0.86	1.48

Table 5.1.4.4**Reliabilities of Domain Scores: Spek S602 Online by IEP Status**

Cluster	Tier	# of Students	# of Tasks	Total Possible Raw Score Points	Cronbach's Alpha	SEM
1	Pre-A	1,881	3	0-6	0.88	0.78
1	A	12,208	6	0-18	0.88	1.38
1	B/C	5,015	6	0-24	0.84	1.63
2-3	Pre-A	3,729	3	0-6	0.87	0.64
2-3	A	20,904	6	0-18	0.84	1.34
2-3	B/C	19,039	6	0-24	0.85	1.56
4-5	Pre-A	818	3	0-6	0.85	0.68
4-5	A	13,564	6	0-18	0.80	1.32
4-5	B/C	30,878	6	0-24	0.85	1.61
6-8	Pre-A	2,091	3	0-6	0.88	0.64
6-8	A	19,109	6	0-18	0.82	1.36
6-8	B/C	38,091	6	0-24	0.86	1.56
9-12	Pre-A	2,437	3	0-6	0.93	0.60
9-12	A	31,289	6	0-18	0.88	1.29
9-12	B/C	25,743	6	0-24	0.89	1.47

5.2 Interrater Agreement Rates

DRC raters score students' responses to the tasks included in the ACCESS Writing and Speaking tests. The scoring of students' responses to these performance tasks is described in Section 4.2. DRC selects a sample of 20% of all responses scored, chosen at random during the operational scoring process, for double scoring. The tables in this section provide information on the interrater agreement rates that the DRC raters achieved. These tables show, for each task, the percentage of agreement between two raters who independently scored students' responses.

For Writing, the first column in the tables shows the task, and the second column shows the number of responses that raters double-scored. The next two columns show the percentages of **agreement** (%AG) and **adjacent agreement** (%AD) that the raters achieved. The last column shows the percentage of **nonadjacent scores** (%NA) that the raters assigned.

The Writing Scoring Scale defines six levels of performance ranging from 0 to 6, with the possibility of awarding a "plus" score between levels (e.g., 3, 3+, or 4 are all valid scores). We considered scores that matched or were contiguous as signifying agreement (%AG)—for example if Rater 1 assigned a score of 3+ while Rater 2 assigned a score of 3, 3+, or 4. We considered scores that were one whole score point apart as adjacent scores (%AD)—for example if Rater 1 assigned a score of 3+ while Rater 2 assigned a score of 2+ or 4+. Finally, if two raters assigned scores that were more than one whole score point apart, we considered

those scores to be nonadjacent scores (%NA). Note that for Writing, DRC reports separate rates of interrater agreement for the raters' scoring of students' keyboarded responses and the raters' scoring of students' handwritten responses.

For Speaking, the first column in the tables shows the task, and the second column shows the number of responses that raters double-scored. The next two columns show the percentages of **exact agreement** (%EX) and **adjacent score agreement** (%AD) that the raters achieved. The last column shows the percentage of **nonadjacent scores** (%NA) that the raters assigned.

The Speaking Scoring Scale defines four levels of performance, ranging from 0 to 4. We considered scores that matched as demonstrating **exact agreement** (%EX). If the scores that two raters assigned differed by one level, we considered those scores to be **adjacent scores** (%AD). Finally, if two raters assigned scores that were more than one level apart, we considered those scores to be **nonadjacent scores** (%NA). Note that the Speaking tasks that target PL 1—the three tasks in the Tier Pre-A forms and the first three tasks in the Tier A forms—are designed for beginning students and use a restricted subset of levels in the Speaking Scoring Scale, with only three possible score levels (see Part 1, Sections 4.2 and 4.4 for more detail). As the range of possible score levels is smaller for these tasks, the rater agreement rates tend to be higher. Therefore, it is not appropriate to compare the interrater agreement rates across tiers, especially when the tasks and the raw score range for the tasks being compared are different.

WIDA stipulates a minimum interrater agreement rate of 70%. For Writing, DRC defines “agreement” as being scored as an adjacent agreement (AG). See Part 1, Section 4.2 for more detail about how WIDA and DRC used the agreement rates to ensure that DRC maintains sufficient quality control throughout scoring.

For Writing, the lowest interrater agreement rate was 91%. For Speaking, the lowest interrater agreement rate was 73%.

5.2.1 Listening

Interrater Agreement is not relevant for the domain of Listening, as all items are multiple-choice items.

5.2.2 Reading

Interrater Agreement is not relevant for the domain of Listening, as all items are multiple-choice items.

5.2.3 Writing

5.2.3.1 Grade 1

Table 5.2.3.1.1

Interrater Agreement: Writ 1 A S602 Online

Task	# in Sample	%AG	%AD	%NA
1	142,918	97	3	0
2	130,144	97	3	0

Table 5.2.3.1.2

Interrater Agreement: Writ 1 B/C S602 Online

Task	# in Sample	%AG	%AD	%NA
1	11,050	91	9	0
2	11,156	97	3	0

5.2.3.2 Grades 2–3

Table 5.2.3.2.1

Interrater Agreement: Writ 2–3 A S602 Online

Task	# in Sample	%AG	%AD	%NA
1	121,848	98	2	0
2	127,844	97	3	0

Table 5.2.3.2.2

Interrater Agreement: Writ 2–3 B/C S602 Online

Task	# in Sample	%AG	%AD	%NA
1	141,680	96	4	0
2	143,086	94	6	0

5.2.3.3 Grades 4–5

Table 5.2.3.3.1

Interrater Agreement: Writ 4–5 A S602 Online

Task	Mode of Response	# in Sample	%AG	%AD	%NA
1	HW	8,144	98	2	0
1	KB	42,642	96	4	0
2	HW	8,062	98	2	0
2	KB	42,436	97	3	0

Table 5.2.3.3.2

Interrater Agreement: Writ 4–5 B/C S602 Online

Task	Mode of Response	# in Sample	%AG	%AD	%NA
1	HW	10,774	97	3	0
1	KB	115,026	97	3	0
2	HW	10,622	96	4	0
2	KB	117,546	97	3	0

5.2.3.4 Grades 6–8

Table 5.2.3.4.1

Interrater Agreement: Writ 6–8 A S602 Online

Task	Mode of Response	# in Sample	%AG	%AD	%NA
1	HW	234	99	1	0
1	KB	88,014	97	3	0
2	HW	214	100	0	0
2	KB	87,808	97	3	0

Table 5.2.3.4.2

Interrater Agreement: Writ 6–8 B/C S602 Online

Task	Mode of Response	# in Sample	%AG	%AD	%NA
1	HW	222	95	5	0
1	KB	116,036	99	1	0
2	HW	198	100	0	0
2	KB	117,366	98	2	0

5.2.3.5 Grades 9–12

Table 5.2.3.5.1

Interrater Agreement: Writ 9–12 A S602 Online

Task	Mode of Response	# in Sample	%AG	%AD	%NA
1	HW	48	100	0	0
1	KB	82,836	97	3	0
2	HW	44	100	0	0
2	KB	83,218	97	3	0

Table 5.2.3.5.2

Interrater Agreement: Writ 9–12 B/C S602 Online

Task	Mode of Response	# in Sample	%AG	%AD	%NA
1	HW	40	95	5	0
1	KB	142,982	99	1	0
2	HW	26	100	0	0
2	KB	153,486	98	2	0

5.2.4 Speaking

5.2.4.1 Grade 1

Table 5.2.4.1.1

Interrater Agreement: Spek 1 Pre-A S602 Online

Task	# in Sample	%EX	%AD	%NA
1	13,638	97	3	0
2	13,210	98	2	0
3	13,610	98	2	0

Table 5.2.4.1.2

Interrater Agreement: Spek 1 A S602 Online

Task	# in Sample	%EX	%AD	%NA
1	73,802	99	1	0
2	73,800	89	10	0
3	72,420	98	2	0
4	72,420	87	13	0
5	74,980	99	1	0
6	74,980	89	11	0

Table 5.2.4.1.3**Interrater Agreement: Spek 1 B/C S602 Online**

Task	# in Sample	%EX	%AD	%NA
1	48,088	84	16	0
2	48,088	86	14	0
3	50,602	79	21	0
4	50,602	78	21	0
5	49,962	85	15	0
6	49,962	80	20	0

5.2.4.2 Grades 2–3**Table 5.2.4.2.1****Interrater Agreement: Spek 2–3 Pre-A S602 Online**

Task	# in Sample	%EX	%AD	%NA
1	19,498	97	3	0
2	18,512	98	2	0
3	18,648	98	2	0

Table 5.2.4.2.2**Interrater Agreement: Spek 2–3 A S602 Online**

Task	# in Sample	%EX	%AD	%NA
1	87,756	99	1	0
2	87,756	83	16	1
3	89,084	99	1	0
4	89,084	84	15	1
5	89,150	99	1	0
6	89,150	84	16	1

Table 5.2.4.2.3**Interrater Agreement: Spek 2–3 B/C S602 Online**

Task	# in Sample	%EX	%AD	%NA
1	138,118	78	22	0
2	138,118	77	22	1
3	141,412	75	24	1
4	141,412	75	24	1
5	138,890	75	24	1
6	138,890	73	25	1

5.2.4.3 Grades 4–5

Table 5.2.4.3.1

Interrater Agreement: Spek 4–5 Pre-A S602 Online

Task	# in Sample	%EX	%AD	%NA
1	10,316	97	3	0
2	10,244	98	2	0
3	10,094	97	3	0

Table 5.2.4.3.2

Interrater Agreement: Spek 4–5 A S602 Online

Task	# in Sample	%EX	%AD	%NA
1	48,550	97	3	0
2	48,550	88	12	0
3	49,008	99	1	0
4	49,006	91	9	0
5	48,338	98	2	0
6	48,344	88	12	0

Table 5.2.4.3.3

Interrater Agreement: Spek 4–5 B/C S602 Online

Task	# in Sample	%EX	%AD	%NA
1	142,628	80	20	0
2	142,628	78	22	0
3	143,872	80	20	0
4	143,872	78	22	0
5	139,934	75	25	0
6	139,932	78	22	0

5.2.4.4 Grades 6–8

Table 5.2.4.4.1

Interrater Agreement: Spek 6–8 Pre-A S602 Online

Task	# in Sample	%EX	%AD	%NA
1	21,070	98	2	0
2	21,480	98	2	0
3	20,530	98	2	0

Table 5.2.4.4.2**Interrater Agreement: Spek 6–8 A S602 Online**

Task	# in Sample	%EX	%AD	%NA
1	65,006	99	1	0
2	65,008	89	11	0
3	68,476	99	1	0
4	68,476	88	12	0
5	68,402	99	1	0
6	68,408	86	13	1

Table 5.2.4.4.3**Interrater Agreement: Spek 6–8 B/C S602 Online**

Task	# in Sample	%EX	%AD	%NA
1	163,442	81	19	0
2	163,440	79	20	0
3	168,846	80	20	0
4	168,846	79	21	0
5	167,642	78	21	1
6	167,648	76	23	1

5.2.4.5 Grades 9–12**Table 5.2.4.5.1****Interrater Agreement: Spek 9–12 Pre-A S602 Online**

Task	# in Sample	%EX	%AD	%NA
1	23,474	99	1	0
2	24,734	98	2	0
3	25,786	98	2	0

Table 5.2.4.5.2**Interrater Agreement: Spek 9–12 A S602 Online**

Task	# in Sample	%EX	%AD	%NA
1	127,068	99	1	0
2	127,044	89	11	0
3	128,402	99	1	0
4	128,402	84	16	1
5	130,748	99	1	0
6	130,740	84	15	0

Table 5.2.4.5.3**Interrater Agreement: Spek 9–12 B/C S602 Online**

Task	# in Sample	%EX	%AD	%NA
1	133,486	81	18	0
2	133,496	80	20	0
3	136,666	77	23	1
4	136,666	75	24	1
5	138,362	80	19	0
6	138,362	78	21	1

5.3 *Conditional Standard Errors of Measurement of the Domain Scale Scores*

The tables in this section present information about the conditional standard errors of measurement (CSEM) values of scale scores at the most important points at which policymakers make decisions such as reclassification about students based on performance on ACCESS—the cut points between language proficiency levels. The CSEM provides information about the amount of measurement error we would expect in any student’s scale score at that point on the underlying latent ability scale. We first computed CSEM values on the theta metric, which is the square root of the inverse of the Test Information Function. Next, we used the multiplicative constant of the linear equation for the domain to linearly transform those logit-based CSEM values so that we could report them on the ACCESS score scale (see Section 2).

When calculated using an IRT approach, CSEM values can vary across the scale scores. For example, in the Listening and Reading domains, if a student answers correctly either a very few or a very large number of items (i.e., scores at the extremes of the scale score distribution), the CSEM value will be larger than it would be if the student correctly answers a moderate number of items. Scale scores near the middle of the score distribution typically have lower CSEM values compared to scale scores near the extremes because many tests are comprised of a large proportion of moderately difficult items, which are well suited to measuring students of moderate proficiency.

We use the CSEM to construct an error band, quantifying the amount of uncertainty in a student’s scale score. One CSEM below a student’s scale score and one CSEM above that scale score indicates an approximate 68% confidence interval. To interpret this confidence interval, consider a student who takes the test 100 times. Assuming measurement error is normally distributed, the student’s true proficiency would fall within the confidence interval 68% of the time (or 68 times out of 100).

As a rule, lower CSEM values around scale scores at important decision points are desirable. Generally speaking, the most important decision points for the ACCESS scores are at the PL 3/4 and PL 4/5 cut points, although the approaches that WIDA states use to make decisions about ACCESS scores differ. As discussed in Section 5, all WIDA states use composite scale

scores when making reclassification decisions, and no WIDA state uses a single domain scale score when making those decisions. Because each grade has its own set of cut points, we provide information for each grade within a grade-level cluster.

Since we scale ACCESS test scores using an IRT approach, CSEM values for the scale scores at the highest cut points are typically large. Use of this approach tends to produce larger CSEM values at the lower and the higher ends of the score scale. In addition, because students exit the EL program when they demonstrate that they are English language proficient, there are typically fewer students at the highest cut points than at those other cut points. Therefore, the CSEM values associated with the scale scores at the highest cut points tend to be larger than those of the scale scores at the lower cut points since there are fewer students available to estimate the scores and the CSEM values for these scores.

Since the Listening and Reading tests are multistage adaptive tests, the CSEM values will vary for the same scale score because the test will route students to take different items; therefore, it is not possible to present a single CSEM value for the scale score that corresponds to each cut point. In the tables for Listening and Reading, the leftmost column shows the proficiency level cut (e.g., 1/2, which is the cut between PL 1 and PL 2). The second column shows the grade level. The third column shows the cut point in the scale score metric (e.g., 305). The next columns present the number of students and the minimum, maximum, mean, and standard deviation of the CSEM values for all students' scale scores at each cut point within a grade level. Note that there are some rare cases where there are no observed scale scores corresponding to certain cut points; therefore, we cannot provide these descriptive statistics. Because Listening and Reading tests are multistage adaptive tests, we would not expect large variation in the mean CSEM values of students' scale scores across cut points within a grade level.

For Writing and Speaking, we present the CSEM values for the scale scores by tier. From these tables, it is possible to determine the extent to which students' responses to the tasks included in the different Writing and Speaking tiers provide targeted information that is useful for accurately placing them into the various proficiency levels. In the tables for Writing and Speaking, the leftmost column shows the proficiency level cut point (e.g., 1/2, which is the cut between PL 1 and PL 2). The second column shows the grade level. The third column shows the cut point in the scale score metric (e.g., 305). In the last column(s), the corresponding CSEM value for the scale score at each cut point are shown.

5.3.1 Listening

5.3.1.1 Grade 1

Table 5.3.1.1

Descriptive Statistics for the Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 1 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	# of Students	Min.	Max.	Mean	Std. Dev.
1/2	1	236	N/A	N/A	N/A	N/A	N/A
2/3	1	259	1,844	15.82	16.33	15.82	0.04
3/4	1	291	226	16.33	17.86	16.57	0.41
4/5	1	303	3,848	16.33	17.35	16.95	0.48
5/6	1	327	254	17.35	18.37	17.55	0.33

5.3.1.2 Grades 2–3

Table 5.3.1.2

Descriptive Statistics for the Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 2–3 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	# of Students	Min.	Max.	Mean	Std. Dev.
1/2	2	245	10	21.43	21.43	21.43	0.00
1/2	3	262	19	22.45	22.45	22.45	0.00
2/3	2	283	106	18.37	18.37	18.37	0.00
2/3	3	300	4	18.37	18.37	18.37	0.00
3/4	2	314	1,215	18.37	19.39	18.68	0.26
3/4	3	331	252	18.37	19.90	19.17	0.55
4/5	2	330	901	18.88	19.39	18.99	0.21
4/5	3	349	819	18.37	19.39	18.91	0.47
5/6	2	354	2,578	18.88	22.96	18.90	0.25
5/6	3	374	N/A	N/A	N/A	N/A	N/A

5.3.1.3 Grades 4–5

Table 5.3.1.3

Descriptive Statistics for the Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 4–5 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	# of Students	Min.	Max.	Mean	Std. Dev.
1/2	4	275	5	17.35	17.86	17.65	0.28
1/2	5	285	121	17.35	19.90	19.77	0.56
2/3	4	313	3	15.82	15.82	15.82	0.00
2/3	5	323	10	15.82	15.82	15.82	0.00
3/4	4	343	N/A	N/A	N/A	N/A	N/A
3/4	5	354	6	17.35	17.35	17.35	0.00
4/5	4	363	293	17.35	17.86	17.36	0.08
4/5	5	375	2,701	17.86	18.37	17.95	0.19
5/6	4	388	400	17.86	18.37	18.21	0.24
5/6	5	401	45	20.41	20.41	20.41	0.00

5.3.1.4 Grades 6–8

Table 5.3.1.4

Descriptive Statistics for the Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 6–8 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	# of Students	Min.	Max.	Mean	Std. Dev.
1/2	6	294	N/A	N/A	N/A	N/A	N/A
1/2	7	302	N/A	N/A	N/A	N/A	N/A
1/2	8	308	N/A	N/A	N/A	N/A	N/A
2/3	6	332	45	16.33	16.33	16.33	0.00
2/3	7	340	4,153	16.33	16.33	16.33	0.00
2/3	8	347	3,302	15.82	16.84	15.82	0.06
3/4	6	363	371	15.82	16.84	16.82	0.11
3/4	7	370	1,377	16.33	16.33	16.33	0.00
3/4	8	377	45	16.33	17.86	16.43	0.39
4/5	6	385	102	16.33	17.35	16.77	0.28
4/5	7	394	1,580	16.84	17.35	16.84	0.04
4/5	8	402	2,098	16.84	17.86	17.30	0.42
5/6	6	411	28	17.86	17.86	17.86	0.00
5/6	7	420	35	18.37	19.90	19.72	0.49
5/6	8	427	10,260	17.86	19.90	17.89	0.23

5.3.1.5 Grades 9–12

Table 5.3.1.5

Descriptive Statistics for the Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 9–12 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	# of Students	Min.	Max.	Mean	Std. Dev.
1/2	9	314	395	20.92	20.92	20.92	0.00
1/2	10	325	N/A	N/A	N/A	N/A	N/A
1/2	11	335	340	19.90	19.90	19.90	0.00
1/2	12	342	59	19.39	19.90	19.72	0.25
2/3	9	353	474	16.84	17.35	16.92	0.19
2/3	10	358	178	16.84	16.84	16.84	0.00
2/3	11	364	20	16.84	16.84	16.84	0.00
2/3	12	368	1,111	16.84	17.35	16.88	0.14
3/4	9	383	66	16.84	17.35	16.84	0.06
3/4	10	389	713	16.84	16.84	16.84	0.00
3/4	11	394	3,105	16.84	16.84	16.84	0.00
3/4	12	398	539	16.84	17.86	17.17	0.24
4/5	9	409	368	16.84	17.35	17.04	0.25
4/5	10	415	1,562	16.84	18.37	17.24	0.26
4/5	11	420	220	16.84	17.86	17.20	0.49
4/5	12	426	101	17.35	18.88	18.01	0.34
5/6	9	434	801	17.35	18.37	17.73	0.49
5/6	10	441	2	18.37	18.37	18.37	0.00
5/6	11	447	20	20.41	20.41	20.41	0.00
5/6	12	452	52	19.90	19.90	19.90	0.00

5.3.2 Reading

5.3.2.1 Grade 1

Table 5.3.2.1

Descriptive Statistics for the Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 1 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	# of Students	Min.	Max.	Mean	Std. Dev.
1/2	1	264	11,323	10.71	12.76	12.13	0.31
2/3	1	286	10,670	9.69	10.71	9.82	0.31
3/4	1	304	4,408	9.69	10.20	10.19	0.09
4/5	1	315	234	9.69	10.20	10.05	0.24
5/6	1	334	N/A	N/A	N/A	N/A	N/A

5.3.2.2 Grades 2–3

Table 5.3.2.2

Descriptive Statistics for the Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 2–3 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	# of Students	Min.	Max.	Mean	Std. Dev.
1/2	2	283	131	11.22	12.24	12.11	0.34
1/2	3	297	96	10.71	10.71	10.71	0.00
2/3	2	307	10,126	10.20	11.22	10.25	0.21
2/3	3	323	5,542	9.69	10.20	9.75	0.16
3/4	2	326	9,247	9.69	10.20	10.20	0.07
3/4	3	342	7,314	9.69	10.20	9.70	0.06
4/5	2	337	143	9.69	10.20	10.01	0.25
4/5	3	352	59	10.20	10.71	10.31	0.21
5/6	2	355	9	10.20	10.20	10.20	0.00
5/6	3	370	1	11.22	11.22	11.22	0.00

5.3.2.3 Grades 4–5

Table 5.3.2.3

Descriptive Statistics for the Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 4–5 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	# of Students	Min.	Max.	Mean	Std. Dev.
1/2	4	307	537	10.71	12.76	12.00	0.48
1/2	5	316	1,851	10.20	12.24	12.00	0.45
2/3	4	335	671	9.69	11.22	10.28	0.22
2/3	5	345	3,792	9.69	10.71	9.83	0.23
3/4	4	354	1,344	9.69	10.71	10.45	0.28
3/4	5	364	5,122	10.20	10.71	10.20	0.01
4/5	4	364	9,810	10.20	10.20	10.20	0.00
4/5	5	373	6,289	10.20	10.71	10.21	0.07
5/6	4	382	55	10.20	10.71	10.68	0.13
5/6	5	391	156	10.71	10.71	10.71	0.00

5.3.2.4 Grades 6–8

Table 5.3.2.4

Descriptive Statistics for the Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 6–8 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	# of Students	Min.	Max.	Mean	Std. Dev.
1/2	6	323	1,073	11.73	12.76	11.84	0.22
1/2	7	329	1,557	11.22	12.24	11.42	0.26
1/2	8	335	87	11.22	11.73	11.50	0.26
2/3	6	353	1,314	10.20	10.71	10.34	0.23
2/3	7	360	272	10.20	10.71	10.22	0.08
2/3	8	366	1,852	10.20	11.22	10.22	0.11
3/4	6	373	1,678	10.20	11.22	10.65	0.17
3/4	7	380	1,024	10.20	11.22	10.43	0.25
3/4	8	386	1,814	10.20	11.73	10.36	0.24
4/5	6	382	2,875	10.20	11.22	10.29	0.19
4/5	7	389	554	10.20	11.22	10.46	0.28
4/5	8	395	3,056	10.20	11.73	10.33	0.35
5/6	6	399	1,055	10.20	11.22	10.21	0.04
5/6	7	406	165	10.71	12.24	10.74	0.18
5/6	8	412	14	11.22	12.76	12.32	0.72

5.3.2.5 Grades 9–12

Table 5.3.2.5

Descriptive Statistics for the Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 9–12 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	# of Students	Min.	Max.	Mean	Std. Dev.
1/2	9	340	3,061	11.22	12.76	11.46	0.29
1/2	10	344	1,940	11.22	12.24	12.00	0.27
1/2	11	348	372	11.22	12.24	11.67	0.51
1/2	12	352	162	11.73	12.76	11.92	0.40
2/3	9	372	419	10.20	10.71	10.24	0.13
2/3	10	377	354	10.20	10.71	10.21	0.05
2/3	11	382	2,188	9.69	10.71	9.90	0.26
2/3	12	386	1,373	9.69	10.71	9.86	0.26
3/4	9	392	829	9.69	10.71	10.17	0.14
3/4	10	397	257	9.69	11.22	10.18	0.15
3/4	11	402	2,043	9.69	11.22	10.20	0.07
3/4	12	407	349	9.69	11.22	10.22	0.14
4/5	9	401	487	9.69	10.71	10.21	0.05
4/5	10	406	3,821	10.20	10.71	10.21	0.06
4/5	11	410	3,190	10.20	11.22	10.22	0.09
4/5	12	414	2,323	10.20	11.73	10.22	0.09
5/6	9	418	3,281	10.20	11.22	10.22	0.08
5/6	10	423	2,987	10.20	11.22	10.22	0.10
5/6	11	427	2,592	10.71	11.73	10.71	0.02
5/6	12	432	1,689	10.71	12.24	10.72	0.04

5.3.3 Writing

5.3.3.1 Grade 1

Table 5.3.3.1

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 1 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	CSEM in Tier A	CSEM in Tier B/C
1/2	1	238	15.84	14.50
2/3	1	275	20.94	19.06
3/4	1	337	20.41	21.48
4/5	1	382	19.87	18.80
5/6	1	405	26.05	20.68

5.3.3.2 Grades 2–3

Table 5.3.3.2

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 2–3 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	CSEM in Tier A	CSEM in Tier B/C
1/2	2	242	14.50	14.23
1/2	3	247	15.04	14.23
2/3	2	279	20.14	19.33
2/3	3	283	20.68	19.87
3/4	2	341	21.21	21.48
3/4	3	346	20.68	21.21
4/5	2	388	18.80	18.53
4/5	3	394	19.33	18.73
5/6	2	411	23.09	21.21
5/6	3	418	25.78	23.36

5.3.3.3 Grades 4–5

Table 5.3.3.3

**Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ
4–5 S602 Online**

Proficiency Level Cut Point	Grade	Cut Score	CSEM in Tier A	CSEM in Tier B/C
1/2	4	266	14.23	19.87
1/2	5	267	14.21	19.60
2/3	4	288	16.92	14.23
2/3	5	293	17.92	14.23
3/4	4	351	21.75	21.48
3/4	5	356	21.75	21.75
4/5	4	401	18.80	20.68
4/5	5	407	18.53	20.14
5/6	4	425	19.60	18.80
5/6	5	433	20.94	18.53

5.3.3.4 Grades 6–8

Table 5.3.3.4

**Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ
6–8 S602 Online**

Proficiency Level Cut Point	Grade	Cut Score	CSEM in Tier A	CSEM in Tier B/C
1/2	6	268	14.77	14.68
1/2	7	273	15.57	14.23
1/2	8	281	17.18	14.50
2/3	6	298	19.87	17.18
2/3	7	305	20.68	18.26
2/3	8	311	21.21	19.33
3/4	6	361	21.21	21.75
3/4	7	367	20.94	21.75
3/4	8	372	20.41	21.48
4/5	6	413	19.06	18.72
4/5	7	419	19.87	18.53
4/5	8	424	20.68	18.53
5/6	6	441	26.05	20.68
5/6	7	450	30.34	23.09
5/6	8	459	35.98	26.31

5.3.3.5 Grades 9–12

Table 5.3.3.5

**Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ
9–12 S602 Online**

Proficiency Level Cut Point	Grade	Cut Score	CSEM in Tier A	CSEM in Tier B/C
1/2	9	289	14.23	14.76
1/2	10	298	14.70	15.00
1/2	11	308	16.38	16.38
1/2	12	318	18.26	17.99
2/3	9	319	18.26	18.26
2/3	10	326	19.33	19.33
2/3	11	335	20.41	20.41
2/3	12	344	21.21	21.21
3/4	9	378	21.75	21.75
3/4	10	385	21.48	21.48
3/4	11	391	21.21	21.21
3/4	12	398	20.79	20.74
4/5	9	430	18.80	18.80
4/5	10	436	18.53	18.80
4/5	11	441	18.80	19.06
4/5	12	447	19.33	19.60
5/6	9	469	24.97	24.97
5/6	10	479	29.27	29.00
5/6	11	490	35.98	35.44
5/6	12	501	44.30	43.50

5.3.4 Speaking

5.3.4.1 Grade 1

Table 5.3.4.1

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 1 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	CSEM in Tier A	CSEM in Tier B/C
1/2	1	205	21.94	16.09
2/3	1	261	27.49	19.89
3/4	1	311	23.98	17.26
4/5	1	361	33.34	21.35
5/6	1	403	59.67	35.68

Note: Tier Pre-A is not presented as it is not possible for Tier Pre-A students to receive a proficiency level higher than 2.

5.3.4.2 Grades 2–3

Table 5.3.4.2

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 2–3 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	CSEM in Tier A	CSEM in Tier B/C
1/2	2	220	22.81	16.38
1/2	3	234	24.86	16.96
2/3	2	273	27.79	19.60
2/3	3	283	26.91	19.51
3/4	2	322	24.28	17.55
3/4	3	332	24.57	17.26
4/5	2	374	31.59	20.22
4/5	3	386	36.27	22.52
5/6	2	415	53.52	31.88
5/6	3	425	62.59	36.56

Note: Tier Pre-A is not presented as it is not possible for Tier Pre-A students to receive a proficiency level higher than 2.

5.3.4.3 Grades 4–5

Table 5.3.4.3

**Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek
4–5 S602 Online**

Proficiency Level Cut Point	Grade	Cut Score	CSEM in Tier A	CSEM in Tier B/C
1/2	4	246	21.64	16.67
1/2	5	258	22.81	16.38
2/3	4	293	27.49	17.84
2/3	5	302	28.08	18.72
3/4	4	342	25.15	19.01
3/4	5	350	24.57	18.43
4/5	4	397	27.20	17.84
4/5	5	407	29.54	18.71
5/6	4	435	40.65	23.11
5/6	5	443	45.04	25.15

Note: Tier Pre-A is not presented as it is not possible for Tier Pre-A students to receive a proficiency level higher than 2.

5.3.4.4 Grades 6–8

Table 5.3.4.4

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 6–8 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	CSEM in Tier A	CSEM in Tier B/C
1/2	6	268	23.69	16.38
1/2	7	277	25.45	17.26
1/2	8	284	26.62	17.84
2/3	6	310	28.37	19.98
2/3	7	317	27.79	20.18
2/3	8	323	27.20	19.89
3/4	6	360	23.98	17.55
3/4	7	369	23.93	17.26
3/4	8	377	24.28	16.96
4/5	6	417	32.17	20.47
4/5	7	425	35.68	22.23
4/5	8	433	39.48	24.28
5/6	6	451	50.60	30.42
5/6	7	457	55.57	33.05
5/6	8	463	61.13	35.97

Note: Tier Pre-A is not presented as it is not possible for Tier Pre-A students to receive a proficiency level higher than 2.

5.3.4.5 Grades 9–12

Table 5.3.4.5

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 9–12 S602 Online

Proficiency Level Cut Point	Grade	Cut Score	CSEM in Tier A	CSEM in Tier B/C
1/2	9	290	25.45	18.13
1/2	10	295	26.32	18.72
1/2	11	299	26.91	19.01
1/2	12	302	27.20	19.30
2/3	9	328	27.49	19.60
2/3	10	333	27.20	19.30
2/3	11	337	26.91	19.30
2/3	12	340	26.62	19.01
3/4	9	385	24.57	17.26
3/4	10	393	24.86	17.55
3/4	11	400	25.74	18.13
3/4	12	406	26.62	18.72
4/5	9	440	36.85	25.45
4/5	10	446	39.78	27.49
4/5	11	451	42.41	29.25
4/5	12	455	44.75	31.00
5/6	9	468	54.11	37.14
5/6	10	471	56.45	38.90
5/6	11	474	58.79	40.65
5/6	12	476	60.84	41.82

Note: Tier Pre-A is not presented as it is not possible for Tier Pre-A students to receive a proficiency level higher than 2.

5.4 Accuracy and Consistency of Domains

One of the main purposes of the WIDA ACCESS program is to identify students' English language proficiency level with respect to the WIDA ELD Standards. Because of the emphasis on classifying student performance, a question of interest is how accurately and consistently ACCESS domain scale scores can classify students into the WIDA proficiency levels determined by the 2016 ACCESS standard-setting process (Cook & MacGregor, 2017). Test users can examine indices that report on the accuracy and consistency of these classifications and can use that information to judge the utility of WIDA's proficiency level categorization, while policymakers can use these indices to assist them when making decisions about ACCESS test design and score reporting (American Educational Research Association et al., 2014). The analyses we conduct to examine the accuracy and consistency of classifications utilize the

methods that Livingston and Lewis (1995) and Young and Yoon (1998) outlined, as implemented in the software program BB-CLASS (Brennan, 2004; cf. also Lee, Hanson, & Brennan, 2002).

Classification accuracy is defined conceptually as the extent to which the proficiency classifications of students based on their observed raw scores or scale scores would agree with those made based on their true scores (Livingston, 2018; Livingston & Lewis, 1995). A student's true score is the average of the scores that the student would have received, averaging over some set of prespecified factors or conditions (e.g., different versions of the test, different times of test administration). Therefore, the calculation of the true scores depends upon the particular factors over which one chooses to average (Livingston, 2018). We assume that true scores measure perfectly, but those scores are unknown. Therefore, to provide the best estimation of classification accuracy for WIDA, we use test data from one ACCESS administration to estimate students' true scale scores based on their domain scale scores and the parameters of the model used in estimating those true scale scores. We can then use the results from our analysis to estimate the percentages of the students who were accurately classified into each proficiency level.

Classification consistency is defined conceptually as the extent to which the proficiency classifications of students agree, given two independent administrations of the same or two parallel test forms. It is impractical to obtain repeated administrations of the same or parallel test forms because of cost, testing burden, and the effects of student memory and practice. However, it is possible to estimate the percentages of the students who would be consistently classified with the assumption that the same test is independently administered twice to the same group of students.

The approach that Livingston and Lewis (1995) took, which we implemented here, uses information about the reliability of the students' domain scale scores, the cut points, and the observed distribution of scores. Then, using a four-parameter beta distribution, we model the distribution of the true scale scores and of the domain scale scores on a parallel form. The Livingston and Lewis procedure requires that the reliability estimate of the students' scores on a test form be provided when calculating the classification consistency and accuracy indices. For Listening and Reading, we used the Rasch student separation reliability estimates by grade-level clusters in the procedure. Since the Writing and Speaking tests were tiered, we needed to produce a single reliability estimate across tiers to implement the Livingston and Lewis procedure. This is a weighted reliability estimate across tiers (see Section 5.1).

Overall classification accuracy indicates the percentage of all students whom we would classify into the same language proficiency level by both their domain scale scores and their true scale scores (i.e., the percentage of students whom we accurately classified). For example, an overall classification accuracy index of 0.774 means that we would classify 77% of the students into the same proficiency level according to their domain scale scores and their true scale scores. **Overall classification consistency** indicates the percentage of all students whom we would classify into the same language proficiency levels by their performances on both the administered test and on a parallel test. For example, an overall classification consistency index of 0.664 means that we would classify 66% of the students into the same

proficiency level if they took two parallel forms of the test. A classification consistency index is always lower than its corresponding classification accuracy index because, in classification consistency, a classification based on a student's performance on the administered test and a classification based on that student's performance on a parallel test are both subject to measurement error. In contrast, in classification accuracy, only the classification based on a student's performance on the administered test contains error while we assume that the classification based on that student's true scale score is free of measurement error.

Overall classification accuracy and consistency indices indicate the degree to which we accurately and consistently classify students into the same WIDA proficiency levels, but not the degree to which we accurately or consistently classify students into the proficiency levels below or above the specific cut point (e.g., at the PL 4/PL 5 cut point). The indices that can address this question are **marginal classification accuracy and consistency indices based on domain scale scores at the cut points**. From an accountability perspective, the most important indices for test users and policymakers to examine are the marginal classification accuracy and consistency indices.

The **marginal classification accuracy indices based on domain scale scores at the cut points** report the percentage of students whom we accurately placed into proficiency levels above and below each cut point based on their domain scale scores. For example, a classification accuracy index of 0.774 at the PL 4/PL 5 cut point means that we would classify 77% of the students in the same way using their domain scale scores or their true scale scores, either into the proficiency levels below the cut point (i.e., PL 1 to PL 4) or into the proficiency levels above the cut point (i.e., PL 5 to PL 6). The **marginal classification consistency indices based on domain scale scores at the cut points** report the percentage of students whom we would classify consistently above and below each cut point based on their domain scale scores. For example, a classification consistency index of 0.664 at the PL 4/PL 5 cut point means that we would classify 66% of the students in the same way if they took two parallel forms, either into the proficiency levels below the cut point (i.e., PL 1 to PL 4) or into the proficiency levels above the cut point (i.e., PL 5 to PL 6). Note that the marginal accuracy and consistency indices are generally higher for students' domain scale scores at the cut points than are the overall classification accuracy and consistency indices (Livingston, 2018). This is because the marginal accuracy and consistency indices report the classification decisions at one cut point at a time while the overall accuracy and consistency indices report the classification decisions at all five cut points at the same time.

The interactions of several factors affect the calculation of classification accuracy and consistency: (1) the number of proficiency level cut points, (2) the magnitude of the test score reliability coefficient, (3) measurement accuracy for scale scores at the cut points, (4) the distances between adjacent cut points, (5) the locations of the cut points on the ability scale, and (6) the proportion of students' scale scores around a cut point (Ercikan & Julian, 2002; Lee et al., 2002). These factors are functions of the test design and, most importantly, the standard-setting decisions. The indices are lower when there is a greater number of proficiency levels, a lower test score reliability coefficient, and higher measurement accuracy of the scale scores at the cut points, as well as when the two adjacent cut points are closer, and when more students' domain scale scores are around a cut point. Furthermore, the numbers and types of

items on a test affect the calculation of the test score reliability coefficient. The lower the test score reliability, the lower the classification accuracy and consistency indices would be. For example, the test score reliability coefficient for the ACCESS Online Writing domain raw scores would be lower than the test score reliability coefficients for similar tests that include more items or tasks since we estimate the test score reliability coefficient for ACCESS Online Writing domain raw scores based on students' performance on only two tasks. Therefore, the classification accuracy and consistency indices for the Writing domain might be lower than those for other domains.

For each test domain, we present three tables. The first reports indices that describe the overall accuracy and overall consistency of the proficiency level classifications for each grade level. The second reports the marginal classification accuracy indices based on domain scale scores at the cut points for each grade level. The third reports the marginal classification consistency indices based on domain scale scores at the cut points for each grade level. If we could not estimate the overall and marginal classification accuracy and consistency indices because we classified fewer than 200 students into a given proficiency level, we combined the affected proficiency level and the proficiency level below it and placed 'N/A' in the table for the affected proficiency level.

Assessment experts have issued little guidance to aid in making judgments about the ideal or expected levels of decision consistency and accuracy needed for educational assessments since many different factors affect the calculation of these indices, as discussed earlier. To help test users and policymakers interpret the results from our classification analyses, for each of the ACCESS test domains, we report the range of the overall classification accuracy and consistency indices across grades. Additionally, we highlight the grade with the lowest classification accuracy and consistency indices. Since the overall accuracy and consistency indices are summaries of the degree of classification accuracy and consistency across all proficiency level cut points, we also report the marginal classification accuracy and consistency indices for these grades to identify the specific source(s) of low classification accuracy and consistency.

For Listening, as shown in Table 5.4.1.1, the overall classification accuracy indices ranged from 0.570 to 0.740, and the overall classification consistency indices ranged from 0.464 to 0.677. Grade 11 had the lowest overall classification accuracy and consistency indices for Listening.

For Reading, as shown in Table 5.4.2.1, the overall classification accuracy indices ranged from 0.589 to 0.708, and the overall classification consistency indices ranged from 0.477 to 0.615. Grade 1 had the lowest overall classification accuracy and consistency indices for Reading.

For Writing, as shown in Table 5.4.3.1, the overall classification accuracy indices ranged from 0.549 to 0.738, and the overall classification consistency indices ranged from 0.498 to 0.640. Grade 5 had the lowest overall classification accuracy and consistency indices for Writing.

For Speaking, as shown in Table 5.4.4.1, the overall classification accuracy indices ranged from 0.626 to 0.767, and the overall classification consistency indices ranged from 0.526 to 0.677. Grade 5 had the lowest overall classification accuracy and consistency indices for Speaking.

From an accountability perspective, the most important indices for test users and policymakers to examine are the marginal classification accuracy and consistency indices. To help them interpret our results, we report for each domain the range of the marginal classification accuracy and consistency indices across grades and then highlight the grades (and the cut points within those grades) that had the lowest marginal classification accuracy and the lowest classification consistency.

For Listening, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.868 to 0.982 (Table 5.4.1.2), and the marginal classification consistency indices ranged from 0.821 to 0.974 (Table 5.4.1.3). Grade 9, at the PL 3/4 cut point, had the lowest marginal classification accuracy and consistency indices.

For Reading, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.853 to 0.982 (Table 5.4.2.2), and the marginal classification consistency indices ranged from 0.801 to 0.972 (Table 5.4.2.3). Grade 1, at the PL 1/2 cut point, had the lowest marginal classification accuracy and consistency indices. Note that grade 1 also had the lowest overall classification accuracy index in the Reading domain. The low marginal classification accuracy and consistency at the PL 1/2 cut point appeared to have contributed to its low overall classification accuracy. However, it should be noted that the marginal classification accuracy and consistency indices for grade 1 Reading are still in the 0.80 to mid-0.90 range.

For Writing, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.654 to 0.998 (Table 5.4.3.2), and the marginal classification consistency indices ranged from 0.631 to 0.998 (Table 5.4.3.3). Grade 5, at the PL 3/4 cut point, had the lowest marginal classification accuracy and consistency indices. Note that grade 5 also had the lowest overall classification accuracy and consistency indices in the Writing domain. For grade 5, the low marginal classification accuracy and consistency at the PL 3/4 cut point appeared to have contributed to their low overall classification accuracy and consistency.

For Speaking, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.806 to 0.998 (Table 5.4.4.2), and the marginal classification consistency indices ranged from 0.766 to 0.998 (Table 5.4.4.3). Grade 7, at the PL 3/4 cut point, had the lowest marginal classification accuracy indices, and grade 12, at the PL 5/6 cut point, has the lowest consistency indices. However, it should be noted that the marginal classification accuracy and consistency indices for grades 7 and 12 Speaking are still in the 0.70 to mid-0.90 range.

When we compared the overall and marginal classification accuracy and consistency indices based on the domain scale scores for a particular grade, we saw that in many instances they told the same story (i.e., for a given grade, when the overall classification accuracy and consistency indices were low, then the marginal classification accuracy and consistency indices also tended to be low).

We observed that in the domains of Listening, Writing, and Speaking, the marginal classification accuracy and consistency indices for PL cut points in the middle of the proficiency level range (i.e., PL 2/3 and PL 3/4 cut points) tended, on average, to be lower than the marginal classification accuracy and consistency indices for cut points at the lower and upper ends of the range, a finding that is consistent with findings from previous researchers (Ercikan & Julian,

2002; Lee et al., 2002). One possible reason might be that the cut points for the proficiency levels in the middle of the proficiency level range tend to be closer together than the cut points for the proficiency levels at the ends of that range. (Cut points tend to be closer to each other when there are many proficiency levels.) We would expect marginal classification accuracy and consistency to vary for different ability levels due to variations in measurement accuracy. That is, the further away the students' domain scale scores are from the cut points, the smaller the classification errors would be, or the more accurate the classification decisions would be. With many proficiency levels, there are more student domain scale scores near the cut points than there would be if there were fewer proficiency levels. Therefore, the higher the number of proficiency levels, the higher the probability that we would misclassify students (Ercikan & Julian, 2002). Additionally, the intervals between cut points that are in the middle of the ACCESS proficiency level range are smaller than the intervals between cut points that are at the upper and lower ends of the proficiency level range. Consequently, the marginal classification accuracy and consistency indices based on the domain scale scores for the PL 2/3 and PL 3/4 cut points tend to be lower than for other cut points, as we might expect.

Although assessment experts have issued little guidance to aid in making judgments about the ideal or expected levels of decision consistency and accuracy needed for educational assessments since many different factors affect the calculation of these indices, as discussed earlier, the ranges of the classification accuracy and consistency indices for the ACCESS domains are very similar to those reported for similar testing programs such as ELPA21 (American Institutes of Research, 2018), except for the Writing domain. Since the ACCESS Online Writing test consists of only two tasks, the test score reliability estimate may be lower than similar writing tests that include more tasks. The classification accuracy and consistency indices derived using the Livingston and Lewis (1995) procedure are affected by the magnitude of the test score reliability, which is lower when a test has fewer tasks. Also note that we would not expect the indices estimated for ACCESS domains to be the same as those computed in other programs, because testing programs differ in their student populations, the numbers of proficiency levels, their test designs, their score distributions, and the methods used to compute classification accuracy and consistency indices. For example, compared to similar testing programs, students taking ACCESS represent a much larger and more diverse population. Additionally, the ACCESS testing program defines more proficiency levels than other similar testing programs, and the ACCESS test design is more complex. Therefore, it is difficult to compare the classification accuracy and consistency indices for ACCESS domains to those for other testing programs.

5.4.1 Listening

Table 5.4.1.1

Overall Accuracy and Consistency of Classification Indices: List S602 Online

Grade	Accuracy	Consistency
1	0.666	0.592
2	0.601	0.505
3	0.599	0.508
4	0.740	0.677
5	0.697	0.627
6	0.612	0.508
7	0.601	0.503
8	0.614	0.524
9	0.575	0.469
10	0.575	0.469
11	0.570	0.464
12	0.577	0.471

Table 5.4.1.2

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: List S602 Online

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.933	0.928	0.914	0.912	0.907
2	0.939	0.907	0.884	0.900	0.922
3	0.938	0.906	0.886	0.901	0.915
4	0.982	0.955	0.932	0.933	0.899
5	0.969	0.951	0.939	0.915	0.879
6	0.975	0.941	0.885	0.884	0.903
7	0.966	0.934	0.887	0.884	0.900
8	0.955	0.928	0.892	0.900	0.901
9	0.952	0.912	0.868	0.889	0.926
10	0.943	0.909	0.873	0.891	0.930
11	0.930	0.902	0.877	0.891	0.937
12	0.926	0.898	0.875	0.906	0.942

Table 5.4.1.3

Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: List S602 Online

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.907	0.896	0.879	0.875	0.870
2	0.916	0.867	0.841	0.857	0.890
3	0.913	0.867	0.843	0.856	0.881
4	0.974	0.938	0.907	0.896	0.857
5	0.958	0.931	0.907	0.877	0.837
6	0.966	0.911	0.842	0.835	0.864
7	0.954	0.902	0.845	0.836	0.860
8	0.939	0.895	0.853	0.855	0.861
9	0.934	0.872	0.821	0.843	0.895
10	0.922	0.868	0.826	0.847	0.900
11	0.902	0.859	0.830	0.849	0.908
12	0.896	0.854	0.829	0.865	0.916

5.4.2 Reading

Table 5.4.2.1

Overall Accuracy and Consistency of Classification Indices: Read S602 Online

Grade	Accuracy	Consistency
1	0.589	0.477
2	0.618	0.507
3	0.607	0.506
4	0.617	0.515
5	0.624	0.525
6	0.708	0.615
7	0.690	0.597
8	0.673	0.580
9	0.662	0.563
10	0.662	0.563
11	0.658	0.562
12	0.672	0.575

Table 5.4.2.2**Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: Read S602 Online**

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.853	0.879	0.916	0.941	0.973
2	0.947	0.901	0.882	0.906	0.962
3	0.915	0.896	0.897	0.911	0.951
4	0.936	0.908	0.899	0.895	0.943
5	0.926	0.906	0.899	0.903	0.948
6	0.906	0.913	0.939	0.950	0.982
7	0.909	0.912	0.927	0.943	0.976
8	0.914	0.909	0.919	0.934	0.968
9	0.929	0.902	0.917	0.927	0.955
10	0.927	0.903	0.920	0.928	0.953
11	0.920	0.907	0.916	0.924	0.951
12	0.918	0.903	0.923	0.931	0.956

Table 5.4.2.3**Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: Read S602 Online**

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.801	0.830	0.883	0.916	0.961
2	0.927	0.859	0.838	0.868	0.944
3	0.881	0.852	0.859	0.877	0.928
4	0.911	0.869	0.858	0.860	0.916
5	0.897	0.866	0.860	0.869	0.924
6	0.869	0.877	0.913	0.932	0.972
7	0.873	0.876	0.899	0.920	0.964
8	0.880	0.873	0.887	0.907	0.952
9	0.902	0.863	0.882	0.898	0.937
10	0.898	0.864	0.885	0.899	0.934
11	0.890	0.870	0.882	0.895	0.930
12	0.886	0.865	0.889	0.902	0.937

5.4.3 Writing

Table 5.4.3.1

Overall Accuracy and Consistency of Classification Indices: Writ S602 Online

Grade	Accuracy	Consistency
1	0.674	0.612
2	0.714	0.611
3	0.724	0.619
4	0.629	0.522
5	0.549	0.498
6	0.738	0.640
7	0.640	0.561
8	0.712	0.596
9	0.621	0.539
10	0.702	0.588
11	0.667	0.570
12	0.677	0.596

Table 5.4.3.2

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: Writ S602 Online

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.888	0.785	0.990	N/A	N/A
2	0.952	0.803	0.951	N/A	N/A
3	0.965	0.897	0.859	0.997	N/A
4	0.963	0.917	0.739	0.990	0.998
5	0.964	0.923	0.654	0.981	N/A
6	0.936	0.884	0.912	N/A	N/A
7	0.929	0.867	0.838	N/A	N/A
8	0.930	0.885	0.889	0.998	N/A
9	0.932	0.877	0.806	0.997	N/A
10	0.941	0.879	0.878	0.996	N/A
11	0.917	0.866	0.876	N/A	N/A
12	0.914	0.885	0.866	N/A	N/A

Table 5.4.3.3**Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: Writ S602 Online**

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.837	0.743	0.989	N/A	N/A
2	0.929	0.731	0.915	N/A	N/A
3	0.949	0.848	0.796	0.996	N/A
4	0.945	0.885	0.654	0.978	0.997
5	0.948	0.894	0.631	0.966	N/A
6	0.907	0.833	0.873	N/A	N/A
7	0.897	0.822	0.816	N/A	N/A
8	0.899	0.836	0.828	0.998	N/A
9	0.901	0.834	0.776	0.995	N/A
10	0.911	0.830	0.823	0.995	N/A
11	0.882	0.815	0.838	N/A	N/A
12	0.880	0.829	0.843	N/A	N/A

5.4.4 Speaking**Table 5.4.4.1****Overall Accuracy and Consistency of Classification Indices: Spek S602 Online**

Grade	Accuracy	Consistency
1	0.724	0.614
2	0.672	0.573
3	0.675	0.543
4	0.638	0.527
5	0.626	0.526
6	0.679	0.583
7	0.648	0.583
8	0.678	0.586
9	0.765	0.677
10	0.767	0.674
11	0.759	0.668
12	0.722	0.641

Table 5.4.4.2

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: Spek S602 Online

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.922	0.871	0.929	0.994	N/A
2	0.928	0.865	0.877	0.988	N/A
3	0.945	0.872	0.857	0.986	0.998
4	0.939	0.883	0.855	0.955	0.994
5	0.932	0.888	0.843	0.952	0.994
6	0.929	0.884	0.864	0.996	N/A
7	0.923	0.893	0.806	0.992	N/A
8	0.927	0.890	0.852	0.995	N/A
9	0.917	0.876	0.965	N/A	N/A
10	0.921	0.869	0.971	N/A	N/A
11	0.923	0.865	0.964	0.997	N/A
12	0.914	0.819	0.982	N/A	N/A

Table 5.4.4.3

Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: Spek S602 Online

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.889	0.819	0.885	0.993	N/A
2	0.896	0.814	0.837	0.987	N/A
3	0.919	0.811	0.776	0.984	0.998
4	0.910	0.841	0.799	0.935	0.994
5	0.902	0.845	0.780	0.940	0.994
6	0.898	0.836	0.820	0.995	N/A
7	0.892	0.846	0.810	0.991	N/A
8	0.895	0.839	0.808	0.994	N/A
9	0.883	0.822	0.940	N/A	N/A
10	0.887	0.810	0.946	N/A	N/A
11	0.889	0.803	0.944	0.997	N/A
12	0.877	0.766	0.970	N/A	N/A

5.5 Reliabilities of Students' Composite Scale Scores

The reliability of the ACCESS composite scale scores indicates the consistency of those scores over replications of the testing procedure. Because the domains that make up the composites consist of different test items, and because items from different domains may measure different abilities (even though items within the domain are assumed to measure a single

ability), a traditional internal consistency index such as Cronbach's coefficient alpha is not appropriate, since statisticians who devised such indices assumed that items in a test measure similar ability. It is more appropriate to report a stratified Cronbach's coefficient alpha (Feldt & Brennan, 1989), which measures consistency in students' composite scale scores when those scores are based on students' responses to sets of items that measure different abilities. A stratified alpha is a weighted average of Cronbach's coefficient alphas for item sets that differ in the maximum score points or "strata." Stratified alpha is a reliability estimate computed by dividing the test into components (strata), computing Cronbach's coefficient alpha separately for the scale scores for each component, and then using the results to estimate a reliability coefficient for the composite scale scores.

In computing the stratified Cronbach's coefficient alphas for ACCESS composite scale scores, we treated each domain that makes up a composite as a separate component (or stratum). For example, when computing the stratified Cronbach's coefficient alphas for students' Literacy scale scores, we entered the variances of the students' scale scores for two components (i.e., Reading and Writing) and the weights of those two components. The stratified Cronbach's coefficient alpha is interpreted like other traditional internal consistency statistics such as Cronbach's coefficient alpha. Like Cronbach's coefficient alpha, a stratified Cronbach's coefficient alpha is an estimate of the proportion of the total variance in the students' composite scale scores that the variance in their true composite scale scores can explain.

Because of the differential weights applied to the ACCESS domains that contribute to the students' composite scale scores, the stratified Cronbach's coefficient alpha is weighted by the contribution that each domain makes to the students' composite scale scores (Kamata, Turhan, & Darandari, 2003; Kane & Case, 2004; Rudner, 2001). Specifically, the formula is

$$\alpha_c = 1 - \frac{\sum_{j=1}^k w_j^2 \sigma_j^2 (1 - \rho_j)}{\sigma_c^2}$$

where

k = the number of components (domains) j that contribute to the composite

w_j = the weight of component (domain) j

σ_j^2 = the variance of the students' scale scores for component (domain) j

σ_c^2 = the variance of the students' composite scale scores

ρ_j = the reliability coefficient for students' scale scores for component (domain) j .

As is true for Cronbach's coefficient alpha (see the explanation in Section 5), there is no one set of criteria that the testing community uses when interpreting stratified Cronbach's coefficient alpha values. There is little consensus among the experts in their views of what the acceptable lower limit of the stratified Cronbach's coefficient alpha value should be, or for that matter, how one should interpret various values. This lack of consensus led the authors of the *Standards for Educational and Psychological Measurement* (2014) to conclude, "The choice of [reliability/precision] estimation and the minimum acceptable level for any index remain a matter of professional judgment" (p. 41).

The tables in this section report the stratified Cronbach's coefficient alphas for the students' scale scores for each of the four composites (Oral, Literacy, Comprehension, and Overall). The first table for each composite provides stratified Cronbach's coefficient alphas for all students' composite scale scores. The second table for each composite provides the same information for the population of female students and the population of male students. The third table provides information by ethnicity, for Hispanic and Other students, and the fourth table provides information for the population of students who have an IEP.

The first column of each table shows the grade-level clusters. The tables report the input values that we used to compute the stratified Cronbach's coefficient alphas (i.e., the number of components for each composite, each component's weight, and the variance of the students' scale scores for each component). See Chapter 3 for an explanation of the procedures we used to compute the composite scale scores.

For the students' scale scores in the Listening and Reading domain components, the reliability coefficient is the Rasch student separation reliability coefficient, provided in Section 5.1.

For the students' scale scores in the Writing and Speaking domain components, which have multiple test forms for each grade-level cluster, we derived a single reliability coefficient for the grade-level cluster. To produce this single value, we weighted Cronbach's coefficient alpha for each of the tiers in the grade-level cluster (provided in Section 5.1) by the number of students who were administered the tier form. The weighted average is shown in the tables.

For each relevant domain component, we report the variance of the students' domain scale scores. We also report the variance of the students' composite scale scores. When we computed the variances of the students' domain scale scores and the variances of the students' composite scale scores, we included the students who had valid scores for all four domains.

Finally, the tables present the computed stratified Cronbach's coefficient alphas for students' scale scores for each composite, by grade-level cluster.

Additionally, we used the stratified Cronbach's coefficient alphas, presented in the tables in this section, to produce the **Accuracy and Consistency** classification tables for the composites (Section 5.7). The stratified Cronbach's coefficient alphas for the Oral scale scores computed for all students was 0.92 (Table 5.5.1.1). The stratified Cronbach's coefficient alphas for the Oral scale scores were 0.92 for male students; ranged from 0.91 to 0.92 for female students (Table 5.5.1.2); 0.92 to 0.93 for Hispanic students; 0.90 to 0.91 for Other students (Table 5.5.1.3); and 0.90 to 0.92 for students with an IEP (Table 5.5.1.4).

The stratified Cronbach's coefficient alphas for the Literacy scale scores computed for all students ranged from 0.89 to 0.90 (Table 5.5.2.1). The stratified Cronbach's coefficient alphas for the Literacy scale scores ranged from 0.89 to 0.90 for male students; 0.88 to 0.90 for female students (Table 5.5.2.2); 0.88 to 0.89 for Hispanic students; 0.87 to 0.90 for Other students (Table 5.5.2.3); and 0.85 to 0.89 for students with an IEP (Table 5.5.2.4).

The stratified Cronbach's coefficient alphas for the Comprehension scale scores computed for all students ranged from 0.91 to 0.93 (Table 5.5.3.1). The stratified Cronbach's coefficient alphas for the Comprehension scale scores ranged from 0.91 to 0.94 for male students; 0.91 to

0.93 for female students (Table 5.5.3.2); 0.89 to 0.93 for Hispanic students; 0.92 to 0.93 for Other students (Table 5.5.3.3); and 0.88 to 0.91 for students with an IEP (Table 5.5.3.4).

Since all WIDA states use students' Overall scale scores in making accountability decisions, the students' Overall scale scores must have high reliability. The stratified Cronbach's coefficient alphas for the Overall scale scores computed for all students was 0.94 (Table 5.5.4.1). The stratified Cronbach's coefficient alphas for the Overall scale scores were 0.94 for male students; ranged from 0.93 to 0.94 for female students (Table 5.5.4.2); was 0.94 for Hispanic students; ranged from 0.93 to 0.94 for Other students (Table 5.5.4.3); and 0.91 to 0.93 for students with an IEP (Table 5.5.4.4).

5.5.1 Oral

Table 5.5.1.1

Reliabilities of Composite Scale Scores: Oral S602 Online

Cluster	Component	Weight	Variance	Reliability
1	Listening	0.50	3525.94	0.89
1	Speaking	0.50	3794.50	0.87
1	Oral	N/A	2937.60	0.92
2-3	Listening	0.50	3286.06	0.88
2-3	Speaking	0.50	4068.78	0.86
2-3	Oral	N/A	3033.22	0.92
4-5	Listening	0.50	3356.60	0.87
4-5	Speaking	0.50	4334.87	0.86
4-5	Oral	N/A	3214.66	0.92
6-8	Listening	0.50	2689.13	0.87
6-8	Speaking	0.50	4215.65	0.87
6-8	Oral	N/A	2831.22	0.92
9-12	Listening	0.50	2541.04	0.86
9-12	Speaking	0.50	3933.68	0.87
9-12	Oral	N/A	2542.93	0.92

Table 5.5.1.2**Reliabilities of Composite Scale Scores: Oral S602 Online by Gender**

Cluster	Component	Weight	Gender	Variance	Reliability
1	Listening	0.50	F	3461.48	0.89
1	Listening	0.50	M	3575.06	0.89
1	Speaking	0.50	F	3842.13	0.87
1	Speaking	0.50	M	3749.75	0.86
1	Oral	N/A	F	2925.89	0.92
1	Oral	N/A	M	2945.60	0.92
2-3	Listening	0.50	F	3114.48	0.87
2-3	Listening	0.50	M	3429.71	0.88
2-3	Speaking	0.50	F	4121.50	0.86
2-3	Speaking	0.50	M	4011.16	0.86
2-3	Oral	N/A	F	2977.33	0.92
2-3	Oral	N/A	M	3080.70	0.92
4-5	Listening	0.50	F	3154.59	0.87
4-5	Listening	0.50	M	3484.50	0.88
4-5	Speaking	0.50	F	4383.54	0.86
4-5	Speaking	0.50	M	4299.26	0.85
4-5	Oral	N/A	F	3136.98	0.92
4-5	Oral	N/A	M	3265.12	0.92
6-8	Listening	0.50	F	2637.09	0.87
6-8	Listening	0.50	M	2698.37	0.87
6-8	Speaking	0.50	F	4316.18	0.87
6-8	Speaking	0.50	M	4109.14	0.86
6-8	Oral	N/A	F	2837.41	0.92
6-8	Oral	N/A	M	2797.16	0.92
9-12	Listening	0.50	F	2455.46	0.86
9-12	Listening	0.50	M	2586.14	0.86
9-12	Speaking	0.50	F	3914.22	0.87
9-12	Speaking	0.50	M	3940.97	0.88
9-12	Oral	N/A	F	2508.45	0.91
9-12	Oral	N/A	M	2557.15	0.92

Table 5.5.1.3**Reliabilities of Composite Scale Scores: Oral S602 Online by Ethnicity**

Cluster	Component	Weight	Ethnicity	Variance	Reliability
1	Listening	0.50	H	3454.59	0.89
1	Listening	0.50	O	3314.61	0.88
1	Speaking	0.50	H	3811.88	0.87
1	Speaking	0.50	O	3339.09	0.85
1	Oral	N/A	H	2915.11	0.93
1	Oral	N/A	O	2588.95	0.91
2-3	Listening	0.50	H	3140.43	0.87
2-3	Listening	0.50	O	3219.73	0.87
2-3	Speaking	0.50	H	4179.98	0.86
2-3	Speaking	0.50	O	3389.65	0.84
2-3	Oral	N/A	H	3019.94	0.92
2-3	Oral	N/A	O	2651.15	0.91
4-5	Listening	0.50	H	3288.47	0.88
4-5	Listening	0.50	O	2959.01	0.85
4-5	Speaking	0.50	H	4366.02	0.86
4-5	Speaking	0.50	O	3482.10	0.84
4-5	Oral	N/A	H	3189.84	0.92
4-5	Oral	N/A	O	2603.08	0.90
6-8	Listening	0.50	H	2635.62	0.87
6-8	Listening	0.50	O	2432.31	0.85
6-8	Speaking	0.50	H	4199.14	0.87
6-8	Speaking	0.50	O	3525.15	0.85
6-8	Oral	N/A	H	2784.14	0.92
6-8	Oral	N/A	O	2398.28	0.91
9-12	Listening	0.50	H	2506.76	0.86
9-12	Listening	0.50	O	2265.82	0.84
9-12	Speaking	0.50	H	3967.52	0.88
9-12	Speaking	0.50	O	3274.78	0.85
9-12	Oral	N/A	H	2520.97	0.92
9-12	Oral	N/A	O	2132.94	0.90

Table 5.5.1.4**Reliabilities of Composite Scale Scores: Oral S602 Online by IEP Status**

Cluster	Component	Weight	Variance	Reliability
1	Listening	0.50	3286.35	0.89
1	Speaking	0.50	3811.36	0.87
1	Oral	N/A	2772.04	0.92
2-3	Listening	0.50	2821.27	0.86
2-3	Speaking	0.50	3663.71	0.85
2-3	Oral	N/A	2546.84	0.91
4-5	Listening	0.50	2412.62	0.85
4-5	Speaking	0.50	3260.36	0.84
4-5	Oral	N/A	2164.06	0.90
6-8	Listening	0.50	1864.96	0.83
6-8	Speaking	0.50	3263.18	0.85
6-8	Oral	N/A	1921.83	0.90
9-12	Listening	0.50	1569.62	0.79
9-12	Speaking	0.50	3508.24	0.88
9-12	Oral	N/A	1763.06	0.90

5.5.2 Literacy**Table 5.5.2.1****Reliabilities of Composite Scale Scores: Litr S602 Online**

Cluster	Component	Weight	Variance	Reliability
1	Reading	0.50	867.57	0.85
1	Writing	0.50	2597.43	0.85
1	Literacy	N/A	1241.13	0.90
2-3	Reading	0.50	1035.30	0.88
2-3	Writing	0.50	3091.55	0.81
2-3	Literacy	N/A	1534.48	0.89
4-5	Reading	0.50	1214.40	0.89
4-5	Writing	0.50	3380.60	0.79
4-5	Literacy	N/A	1840.65	0.89
6-8	Reading	0.50	1159.06	0.89
6-8	Writing	0.50	2093.91	0.79
6-8	Literacy	N/A	1355.76	0.90
9-12	Reading	0.50	1284.54	0.90
9-12	Writing	0.50	1770.27	0.77
9-12	Literacy	N/A	1183.05	0.89

Table 5.5.2.2**Reliabilities of Composite Scale Scores: LitR S602 Online by Gender**

Cluster	Component	Weight	Gender	Variance	Reliability
1	Reading	0.50	F	891.66	0.86
1	Reading	0.50	M	856.85	0.85
1	Writing	0.50	F	2520.33	0.84
1	Writing	0.50	M	2678.86	0.85
1	Literacy	N/A	F	1240.77	0.90
1	Literacy	N/A	M	1256.30	0.90
2-3	Reading	0.50	F	1015.03	0.88
2-3	Reading	0.50	M	1049.94	0.88
2-3	Writing	0.50	F	3034.71	0.80
2-3	Writing	0.50	M	3109.68	0.82
2-3	Literacy	N/A	F	1513.87	0.88
2-3	Literacy	N/A	M	1543.01	0.89
4-5	Reading	0.50	F	1144.87	0.89
4-5	Reading	0.50	M	1245.48	0.90
4-5	Writing	0.50	F	3279.69	0.77
4-5	Writing	0.50	M	3448.60	0.80
4-5	Literacy	N/A	F	1779.58	0.88
4-5	Literacy	N/A	M	1872.59	0.89
6-8	Reading	0.50	F	1124.42	0.88
6-8	Reading	0.50	M	1172.64	0.89
6-8	Writing	0.50	F	2121.96	0.78
6-8	Writing	0.50	M	2065.67	0.80
6-8	Literacy	N/A	F	1355.74	0.89
6-8	Literacy	N/A	M	1346.24	0.90
9-12	Reading	0.50	F	1227.89	0.90
9-12	Reading	0.50	M	1307.52	0.90
9-12	Writing	0.50	F	1737.68	0.76
9-12	Writing	0.50	M	1795.07	0.78
9-12	Literacy	N/A	F	1156.17	0.88
9-12	Literacy	N/A	M	1193.59	0.89

Table 5.5.2.3**Reliabilities of Composite Scale Scores: LitR S602 Online by Ethnicity**

Cluster	Component	Weight	Ethnicity	Variance	Reliability
1	Reading	0.50	H	700.65	0.82
1	Reading	0.50	O	1105.49	0.88
1	Writing	0.50	H	2513.44	0.86
1	Writing	0.50	O	2209.20	0.82
1	Literacy	N/A	H	1083.69	0.89
1	Literacy	N/A	O	1282.45	0.90
2-3	Reading	0.50	H	960.83	0.87
2-3	Reading	0.50	O	1101.54	0.89
2-3	Writing	0.50	H	3238.82	0.83
2-3	Writing	0.50	O	2240.30	0.77
2-3	Literacy	N/A	H	1525.68	0.89
2-3	Literacy	N/A	O	1292.09	0.87
4-5	Reading	0.50	H	1160.97	0.89
4-5	Reading	0.50	O	1215.60	0.90
4-5	Writing	0.50	H	3424.16	0.79
4-5	Writing	0.50	O	2524.58	0.77
4-5	Literacy	N/A	H	1821.74	0.89
4-5	Literacy	N/A	O	1523.05	0.88
6-8	Reading	0.50	H	1106.65	0.88
6-8	Reading	0.50	O	1195.97	0.89
6-8	Writing	0.50	H	2073.38	0.79
6-8	Writing	0.50	O	1735.46	0.76
6-8	Literacy	N/A	H	1315.47	0.89
6-8	Literacy	N/A	O	1234.35	0.89
9-12	Reading	0.50	H	1209.52	0.89
9-12	Reading	0.50	O	1394.06	0.91
9-12	Writing	0.50	H	1748.88	0.78
9-12	Writing	0.50	O	1549.43	0.74
9-12	Literacy	N/A	H	1136.69	0.88
9-12	Literacy	N/A	O	1135.42	0.88

Table 5.5.2.4**Reliabilities of Composite Scale Scores: Litr S602 Online by IEP Status**

Cluster	Component	Weight	Variance	Reliability
1	Reading	0.50	660.23	0.80
1	Writing	0.50	2676.45	0.86
1	Literacy	N/A	1089.66	0.89
2-3	Reading	0.50	820.25	0.84
2-3	Writing	0.50	2761.40	0.85
2-3	Literacy	N/A	1210.54	0.89
4-5	Reading	0.50	979.63	0.86
4-5	Writing	0.50	2589.82	0.82
4-5	Literacy	N/A	1350.83	0.89
6-8	Reading	0.50	835.99	0.84
6-8	Writing	0.50	1366.26	0.79
6-8	Literacy	N/A	863.18	0.88
9-12	Reading	0.50	949.53	0.87
9-12	Writing	0.50	1296.64	0.76
9-12	Literacy	N/A	746.69	0.85

5.5.3 Comprehension**Table 5.5.3.1****Reliabilities of Composite Scale Scores: Cphn S602 Online**

Cluster	Component	Weight	Variance	Reliability
1	Listening	0.30	3525.94	0.89
1	Reading	0.70	867.57	0.85
1	Comprehension	N/A	1048.99	0.91
2-3	Listening	0.30	3286.06	0.88
2-3	Reading	0.70	1035.30	0.88
2-3	Comprehension	N/A	1264.54	0.92
4-5	Listening	0.30	3356.60	0.87
4-5	Reading	0.70	1214.40	0.89
4-5	Comprehension	N/A	1490.43	0.93
6-8	Listening	0.30	2689.13	0.87
6-8	Reading	0.70	1159.06	0.89
6-8	Comprehension	N/A	1316.03	0.93
9-12	Listening	0.30	2541.04	0.86
9-12	Reading	0.70	1284.54	0.90
9-12	Comprehension	N/A	1375.84	0.93

Table 5.5.3.2**Reliabilities of Composite Scale Scores: Cphn S602 Online by Gender**

Cluster	Component	Weight	Gender	Variance	Reliability
1	Listening	0.30	F	3461.48	0.89
1	Listening	0.30	M	3575.06	0.89
1	Reading	0.70	F	891.66	0.86
1	Reading	0.70	M	856.85	0.85
1	Comprehension	N/A	F	1059.22	0.91
1	Comprehension	N/A	M	1048.42	0.91
2-3	Listening	0.30	F	3114.48	0.87
2-3	Listening	0.30	M	3429.71	0.88
2-3	Reading	0.70	F	1015.03	0.88
2-3	Reading	0.70	M	1049.94	0.88
2-3	Comprehension	N/A	F	1222.53	0.92
2-3	Comprehension	N/A	M	1298.59	0.92
4-5	Listening	0.30	F	3154.59	0.87
4-5	Listening	0.30	M	3484.50	0.88
4-5	Reading	0.70	F	1144.87	0.89
4-5	Reading	0.70	M	1245.48	0.90
4-5	Comprehension	N/A	F	1402.40	0.93
4-5	Comprehension	N/A	M	1534.67	0.94
6-8	Listening	0.30	F	2637.09	0.87
6-8	Listening	0.30	M	2698.37	0.87
6-8	Reading	0.70	F	1124.42	0.88
6-8	Reading	0.70	M	1172.64	0.89
6-8	Comprehension	N/A	F	1285.57	0.92
6-8	Comprehension	N/A	M	1324.22	0.93
9-12	Listening	0.30	F	2455.46	0.86
9-12	Listening	0.30	M	2586.14	0.86
9-12	Reading	0.70	F	1227.89	0.90
9-12	Reading	0.70	M	1307.52	0.90
9-12	Comprehension	N/A	F	1328.74	0.93
9-12	Comprehension	N/A	M	1394.83	0.93

Table 5.5.3.3**Reliabilities of Composite Scale Scores: Cphn S602 Online by Ethnicity**

Cluster	Component	Weight	Ethnicity	Variance	Reliability
1	Listening	0.30	H	3454.59	0.89
1	Listening	0.30	O	3314.61	0.88
1	Reading	0.70	H	700.65	0.82
1	Reading	0.70	O	1105.49	0.88
1	Comprehension	N/A	H	883.26	0.89
1	Comprehension	N/A	O	1231.38	0.92
2-3	Listening	0.30	H	3140.43	0.87
2-3	Listening	0.30	O	3219.73	0.87
2-3	Reading	0.70	H	960.83	0.87
2-3	Reading	0.70	O	1101.54	0.89
2-3	Comprehension	N/A	H	1160.18	0.92
2-3	Comprehension	N/A	O	1333.45	0.93
4-5	Listening	0.30	H	3288.47	0.88
4-5	Listening	0.30	O	2959.01	0.85
4-5	Reading	0.70	H	1160.97	0.89
4-5	Reading	0.70	O	1215.60	0.90
4-5	Comprehension	N/A	H	1427.49	0.93
4-5	Comprehension	N/A	O	1420.09	0.93
6-8	Listening	0.30	H	2635.62	0.87
6-8	Listening	0.30	O	2432.31	0.85
6-8	Reading	0.70	H	1106.65	0.88
6-8	Reading	0.70	O	1195.97	0.89
6-8	Comprehension	N/A	H	1258.32	0.92
6-8	Comprehension	N/A	O	1300.05	0.93
9-12	Listening	0.30	H	2506.76	0.86
9-12	Listening	0.30	O	2265.82	0.84
9-12	Reading	0.70	H	1209.52	0.89
9-12	Reading	0.70	O	1394.06	0.91
9-12	Comprehension	N/A	H	1306.59	0.93
9-12	Comprehension	N/A	O	1408.17	0.93

Table 5.5.3.4**Reliabilities of Composite Scale Scores: Cphn S602 Online by IEP Status**

Cluster	Component	Weight	Variance	Reliability
1	Listening	0.30	3286.35	0.89
1	Reading	0.70	660.23	0.80
1	Comprehension	N/A	809.53	0.88
2-3	Listening	0.30	2821.27	0.86
2-3	Reading	0.70	820.25	0.84
2-3	Comprehension	N/A	926.27	0.89
4-5	Listening	0.30	2412.62	0.85
4-5	Reading	0.70	979.63	0.86
4-5	Comprehension	N/A	1054.37	0.91
6-8	Listening	0.30	1864.96	0.83
6-8	Reading	0.70	835.99	0.84
6-8	Comprehension	N/A	866.04	0.89
9-12	Listening	0.30	1569.62	0.79
9-12	Reading	0.70	949.53	0.87
9-12	Comprehension	N/A	874.77	0.90

5.5.4 Overall

Table 5.5.4.1

Reliabilities of Composite Scale Scores: Over S602 Online

Cluster	Component	Weight	Variance	Reliability
1	Listening	0.15	3525.94	0.89
1	Reading	0.35	867.57	0.85
1	Writing	0.35	2597.43	0.85
1	Speaking	0.15	3794.50	0.87
1	Overall Composite	N/A	1375.82	0.94
2-3	Listening	0.15	3286.06	0.88
2-3	Reading	0.35	1035.30	0.88
2-3	Writing	0.35	3091.55	0.81
2-3	Speaking	0.15	4068.78	0.86
2-3	Overall Composite	N/A	1700.10	0.94
4-5	Listening	0.15	3356.60	0.87
4-5	Reading	0.35	1214.40	0.89
4-5	Writing	0.35	3380.60	0.79
4-5	Speaking	0.15	4334.87	0.86
4-5	Overall Composite	N/A	1993.92	0.94
6-8	Listening	0.15	2689.13	0.87
6-8	Reading	0.35	1159.06	0.89
6-8	Writing	0.35	2093.91	0.79
6-8	Speaking	0.15	4215.65	0.87
6-8	Overall Composite	N/A	1563.33	0.94
9-12	Listening	0.15	2541.04	0.86
9-12	Reading	0.35	1284.54	0.90
9-12	Writing	0.35	1770.27	0.77
9-12	Speaking	0.15	3933.68	0.87
9-12	Overall Composite	N/A	1373.73	0.94

Table 5.5.4.2**Reliabilities of Composite Scale Scores: Over S602 Online by Gender**

Cluster	Component	Weight	Gender	Variance	Reliability
1	Listening	0.15	F	3461.48	0.89
1	Listening	0.15	M	3575.06	0.89
1	Reading	0.35	F	891.66	0.86
1	Reading	0.35	M	856.85	0.85
1	Writing	0.35	F	2520.33	0.84
1	Writing	0.35	M	2678.86	0.85
1	Speaking	0.15	F	3842.13	0.87
1	Speaking	0.15	M	3749.75	0.86
1	Overall Composite	N/A	F	1372.21	0.94
1	Overall Composite	N/A	M	1387.49	0.94
2	Listening	0.15	F	3114.48	0.87
2	Listening	0.15	M	3429.71	0.88
2	Reading	0.35	F	1015.03	0.88
2	Reading	0.35	M	1049.94	0.88
2	Writing	0.35	F	3034.71	0.80
2	Writing	0.35	M	3109.68	0.82
2	Speaking	0.15	F	4121.50	0.86
2	Speaking	0.15	M	4011.16	0.86
2	Overall Composite	N/A	F	1675.95	0.93
2	Overall Composite	N/A	M	1713.27	0.94
4-5	Listening	0.15	F	3154.59	0.87
4-5	Listening	0.15	M	3484.50	0.88
4-5	Reading	0.35	F	1144.87	0.89
4-5	Reading	0.35	M	1245.48	0.90
4-5	Writing	0.35	F	3279.69	0.77
4-5	Writing	0.35	M	3448.60	0.80
4-5	Speaking	0.15	F	4383.54	0.86
4-5	Speaking	0.15	M	4299.26	0.85
4-5	Overall Composite	N/A	F	1935.61	0.93
4-5	Overall Composite	N/A	M	2025.47	0.94
6-8	Listening	0.15	F	2637.09	0.87
6-8	Listening	0.15	M	2698.37	0.87
6-8	Reading	0.35	F	1124.42	0.88
6-8	Reading	0.35	M	1172.64	0.89
6-8	Writing	0.35	F	2121.96	0.78
6-8	Writing	0.35	M	2065.67	0.80
6-8	Speaking	0.15	F	4316.18	0.87

Cluster	Component	Weight	Gender	Variance	Reliability
6-8	Speaking	0.15	M	4109.14	0.86
6-8	Overall Composite	N/A	F	1570.04	0.94
6-8	Overall Composite	N/A	M	1546.09	0.94
9-12	Listening	0.15	F	2455.46	0.86
9-12	Listening	0.15	M	2586.14	0.86
9-12	Reading	0.35	F	1227.89	0.90
9-12	Reading	0.35	M	1307.52	0.90
9-12	Writing	0.35	F	1737.68	0.76
9-12	Writing	0.35	M	1795.07	0.78
9-12	Speaking	0.15	F	3914.22	0.87
9-12	Speaking	0.15	M	3940.97	0.88
9-12	Overall Composite	N/A	F	1351.60	0.94
9-12	Overall Composite	N/A	M	1380.63	0.94

Table 5.5.4.3**Reliabilities of Composite Scale Scores: Over S602 Online by Ethnicity**

Cluster	Component	Weight	Ethnicity	Variance	Reliability
1	Listening	0.15	H	3454.59	0.89
1	Listening	0.15	O	3314.61	0.88
1	Reading	0.35	H	700.65	0.82
1	Reading	0.35	O	1105.49	0.88
1	Writing	0.35	H	2513.44	0.86
1	Writing	0.35	O	2209.20	0.82
1	Speaking	0.15	H	3811.88	0.87
1	Speaking	0.15	O	3339.09	0.85
1	Overall Composite	N/A	H	1243.49	0.94
1	Overall Composite	N/A	O	1339.72	0.94
2	Listening	0.15	H	3140.43	0.87
2	Listening	0.15	O	3219.73	0.87
2	Reading	0.35	H	960.83	0.87
2	Reading	0.35	O	1101.54	0.89
2	Writing	0.35	H	3238.82	0.83
2	Writing	0.35	O	2240.30	0.77
2	Speaking	0.15	H	4179.98	0.86
2	Speaking	0.15	O	3389.65	0.84
2	Overall Composite	N/A	H	1682.53	0.94
2	Overall Composite	N/A	O	1437.09	0.93
4-5	Listening	0.15	H	3288.47	0.88
4-5	Listening	0.15	O	2959.01	0.85
4-5	Reading	0.35	H	1160.97	0.89
4-5	Reading	0.35	O	1215.60	0.90
4-5	Writing	0.35	H	3424.16	0.79
4-5	Writing	0.35	O	2524.58	0.77
4-5	Speaking	0.15	H	4366.02	0.86
4-5	Speaking	0.15	O	3482.10	0.84
4-5	Overall Composite	N/A	H	1969.93	0.94
4-5	Overall Composite	N/A	O	1607.95	0.93
6-8	Listening	0.15	H	2635.62	0.87
6-8	Listening	0.15	O	2432.31	0.85
6-8	Reading	0.35	H	1106.65	0.88
6-8	Reading	0.35	O	1195.97	0.89
6-8	Writing	0.35	H	2073.38	0.79
6-8	Writing	0.35	O	1735.46	0.76
6-8	Speaking	0.15	H	4199.14	0.87

Cluster	Component	Weight	Ethnicity	Variance	Reliability
6-8	Speaking	0.15	O	3525.15	0.85
6-8	Overall Composite	N/A	H	1517.09	0.94
6-8	Overall Composite	N/A	O	1372.14	0.94
9-12	Listening	0.15	H	2506.76	0.86
9-12	Listening	0.15	O	2265.82	0.84
9-12	Reading	0.35	H	1209.52	0.89
9-12	Reading	0.35	O	1394.06	0.91
9-12	Writing	0.35	H	1748.88	0.78
9-12	Writing	0.35	O	1549.43	0.74
9-12	Speaking	0.15	H	3967.52	0.88
9-12	Speaking	0.15	O	3274.78	0.85
9-12	Overall Composite	N/A	H	1330.38	0.94
9-12	Overall Composite	N/A	O	1243.55	0.93

Table 5.5.4.4**Reliabilities of Composite Scale Scores: Over S602 Online by IEP Status**

Cluster	Component	Weight	Variance	Reliability
1	Listening	0.15	3286.35	0.89
1	Reading	0.35	660.23	0.80
1	Writing	0.35	2676.45	0.86
1	Speaking	0.15	3811.36	0.87
1	Overall Composite	N/A	1169.43	0.93
2-3	Listening	0.15	2821.27	0.86
2-3	Reading	0.35	820.25	0.84
2-3	Writing	0.35	2761.40	0.85
2-3	Speaking	0.15	3663.71	0.85
2-3	Overall Composite	N/A	1267.82	0.93
4-5	Listening	0.15	2412.62	0.85
4-5	Reading	0.35	979.63	0.86
4-5	Writing	0.35	2589.82	0.82
4-5	Speaking	0.15	3260.36	0.84
4-5	Overall Composite	N/A	1305.76	0.93
6-8	Listening	0.15	1864.96	0.83
6-8	Reading	0.35	835.99	0.84
6-8	Writing	0.35	1366.26	0.79
6-8	Speaking	0.15	3263.18	0.85
6-8	Overall Composite	N/A	932.35	0.92
9-12	Listening	0.15	1569.62	0.79
9-12	Reading	0.35	949.53	0.87
9-12	Writing	0.35	1296.64	0.76
9-12	Speaking	0.15	3508.24	0.88
9-12	Overall Composite	N/A	812.10	0.91

5.6 Conditional Standard Errors of Measurement of the Composite Scale Scores

CSEMs for the four ACCESS composite scale scores provide test users with a benchmark indicating how free a student's composite scale score is from measurement errors at different WIDA proficiency levels. Due to the differential weights applied to different ACCESS domains (see the introduction to Section 3 for weighting conventions), WIDA estimates the CSEMs using a procedure that is based on IRT (Lord, 1980) and developed by Price, Lurie, Raju, Wilkins, and Zhu (2006). Price et al. (2006) extended the work by Lord (1980) and Kolen, Hanson, and Brennan (1992) in estimating the CSEMs of students' composite scale scores consisting of components. The basic premise of this procedure is that one can empirically estimate the CSEM for a student's weighted composite scale score using the IRT-based CSEMs for each student's component scale scores and the weights associated with the components. We used this method to estimate the CSEMs for ACCESS composite scale scores by treating the ACCESS domains as components.

We used a three-step process to derive the CSEM for each ACCESS composite scale score. We calculated a unique CSEM for each composite scale score by grade. Since this procedure relies on empirical student data, which are subject to year-to-year fluctuations, we used all population student data from all previous three ACCESS 2.0 series in our calculations to obtain more stable estimates than using data from just a single series.

Step 1. Since we calibrated ACCESS domains separately, measurement errors associated with each of the ACCESS domains, as expressed in the CSEM, were independent of each other. Therefore, we estimated the CSEM for a student's composite scale score x , SEM_x , using the equation derived by Price et al. (2006):

$$SEM_x = \sqrt{W_1^2 SEM_1^2 + W_2^2 SEM_2^2 + W_3^2 SEM_3^2 + \dots + W_k^2 SEM_k^2}$$

Where SEM_i^2 is the student's IRT-based score error variance or the squared CSEM for the student's scale score for ACCESS domain i , and W_i is the weight applied to domain i , for $i=1,\dots,k$.

Step 2. Due to the differential weights applied to different ACCESS domains, two students with the same weighted domain scale scores may have composite scale scores with different CSEMs; therefore, we instituted an additional step to obtain a unique CSEM value for each composite scale score. Specifically, we estimated the expected value of the CSEM functions for a composite scale score using a regression approach, and we reported this expected value as the CSEM for that composite scale score.

Step 3. We applied a linear smoothing procedure to derive the CSEMs for composite scale scores that we did not observe in the data.

The figures in this section show graphically the CSEMs for various composite scale scores by grade level. The students' composite scale scores appear on the horizontal axis, and the corresponding CSEMs appear on the vertical axis. Each point in a figure represents a student in the dataset, showing the relationship between the CSEM and that student's composite scale score. We did not plot values for students who received the lowest possible scale scores for any

ACCESS domains, as it is not possible to compute accurately the CSEM for these students' scale scores. For grade-level clusters with multiple grades, we use different colors in the figures to represent students in different grades.

The five vertical lines in the figure indicate the five ACCESS composite scale score cut points for the highest grade in the grade-level cluster for the test form, dividing the figure into six sections representing the six WIDA proficiency levels.

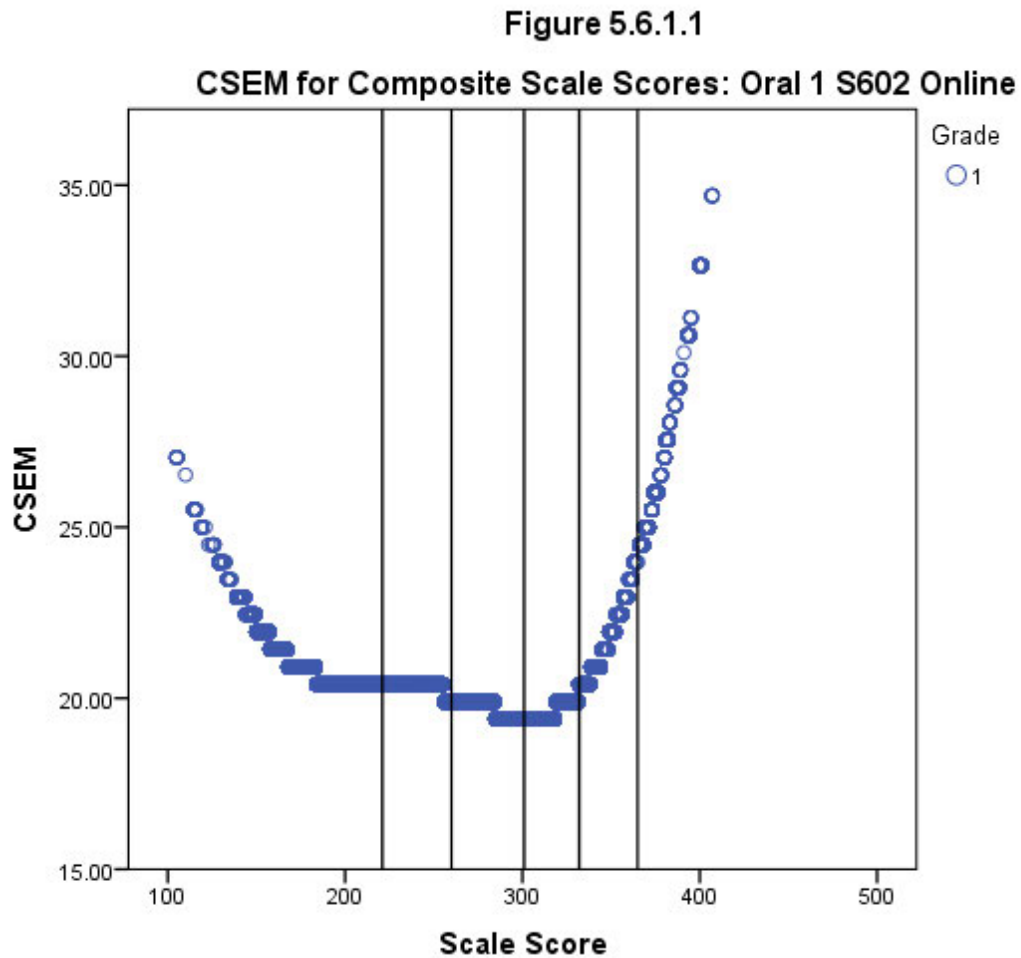
Smaller CSEM values indicate less measurement error (i.e., greater measurement accuracy). In general, these figures show that the CSEMs are smaller and fairly constant in the middle of the composite scale score range but larger and more variable for extremely low and high composite scale scores. This is to be expected since we used an IRT approach when scaling ACCESS, which typically produces larger CSEMs for scale scores that are at the lower and the higher ends of the scale score range. In addition, because students exit the EL program when they demonstrate that they are English language proficient, the number of students whose composite scale scores are at the extreme high end of the score range is typically small, as compared to the number of students whose composite scale scores are in the middle of the score range. Therefore, the measurement errors associated with the composite scale scores at the extremely high end of the score range tend to be larger since the calculation of these scale scores is based on the test performances of fewer students.

5.6.1 Oral

5.6.1.1 Grade 1

Figure 5.6.1.1

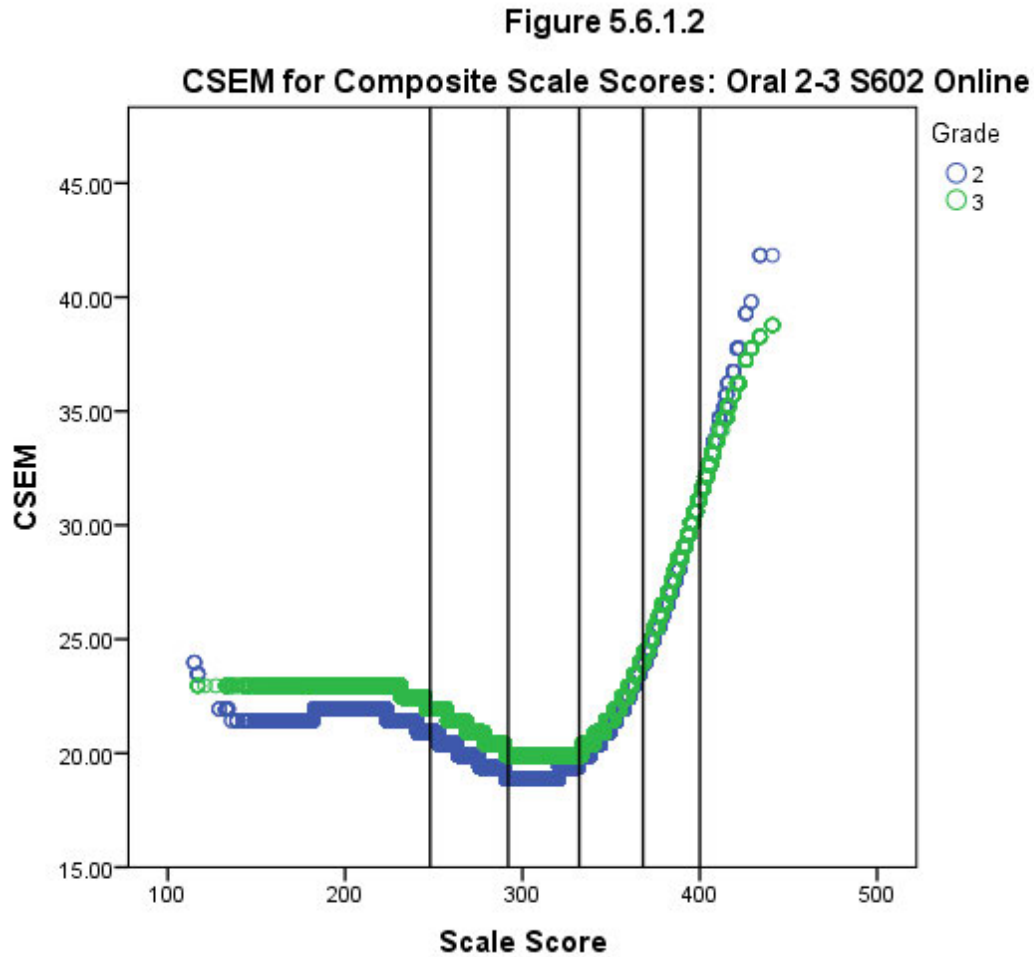
CSEM for Composite Scale Scores: Oral 1 S602 Online



5.6.1.2 Grades 2–3

Figure 5.6.1.2

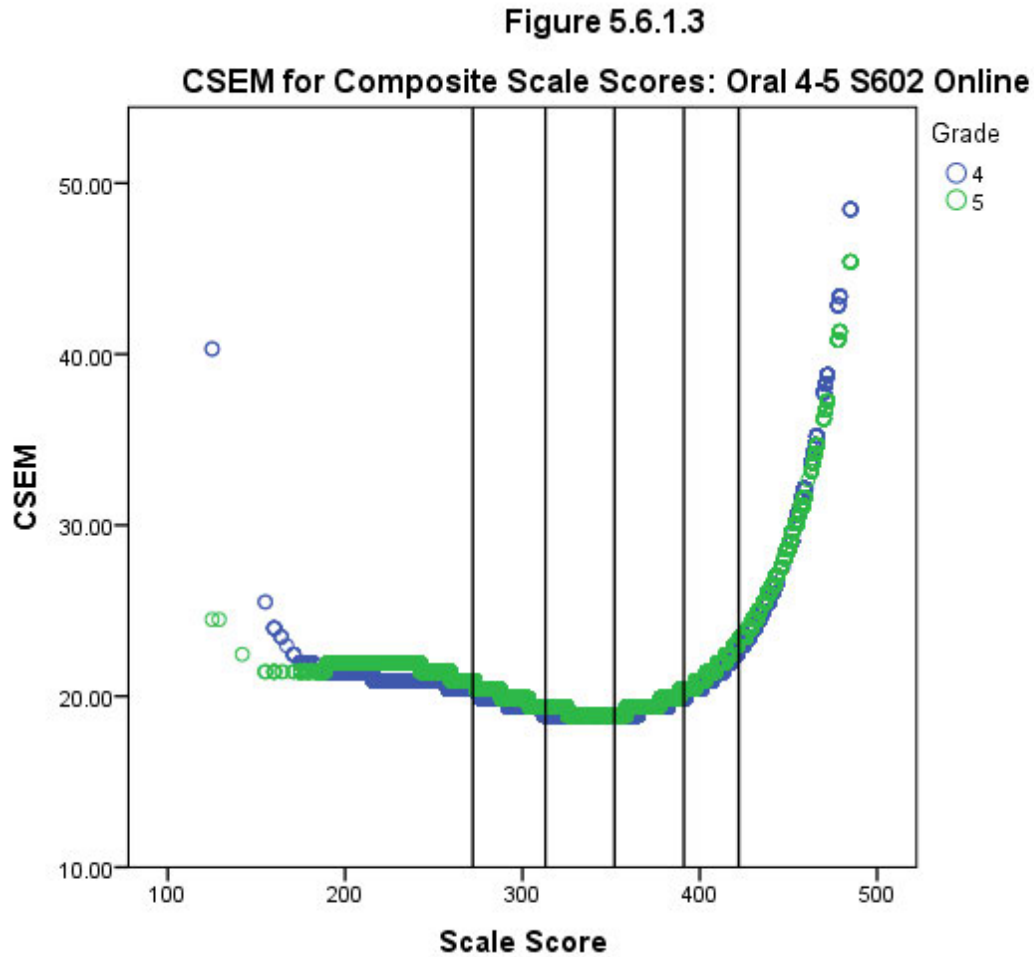
CSEM for Composite Scale Scores: Oral 2–3 S602 Online



5.6.1.3 Grades 4–5

Figure 5.6.1.3

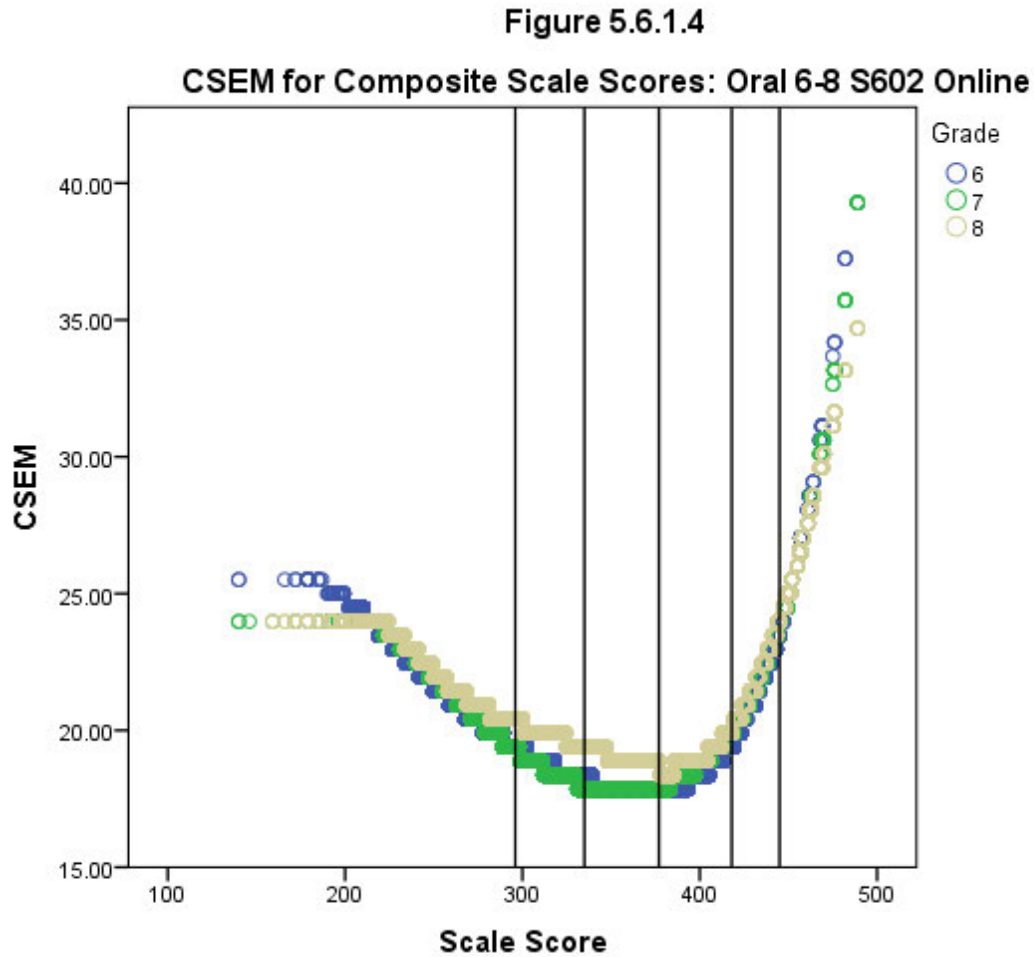
CSEM for Composite Scale Scores: Oral 4–5 S602 Online



5.6.1.4 Grades 6–8

Figure 5.6.1.4

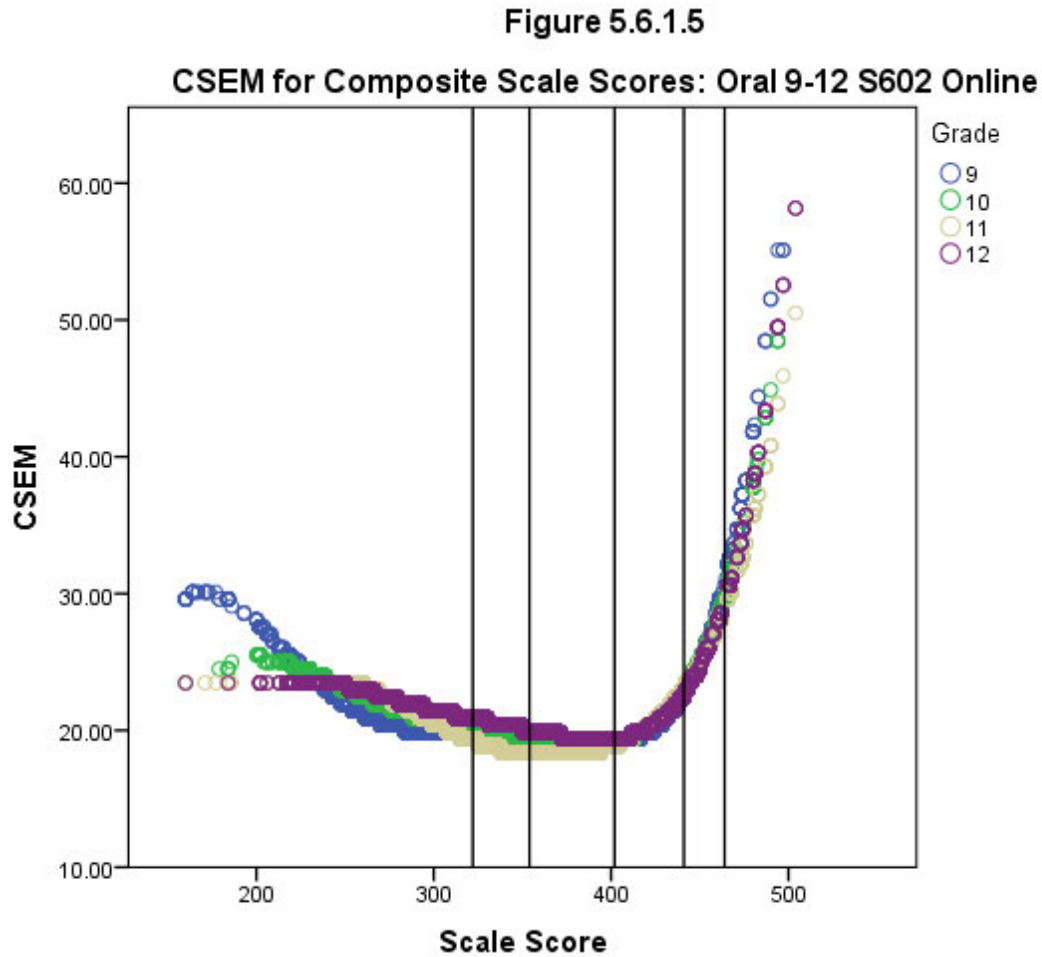
CSEM for Composite Scale Scores: Oral 6–8 S602 Online



5.6.1.5 Grades 9–12

Figure 5.6.1.5

CSEM for Composite Scale Scores: Oral 9–12 S602 Online

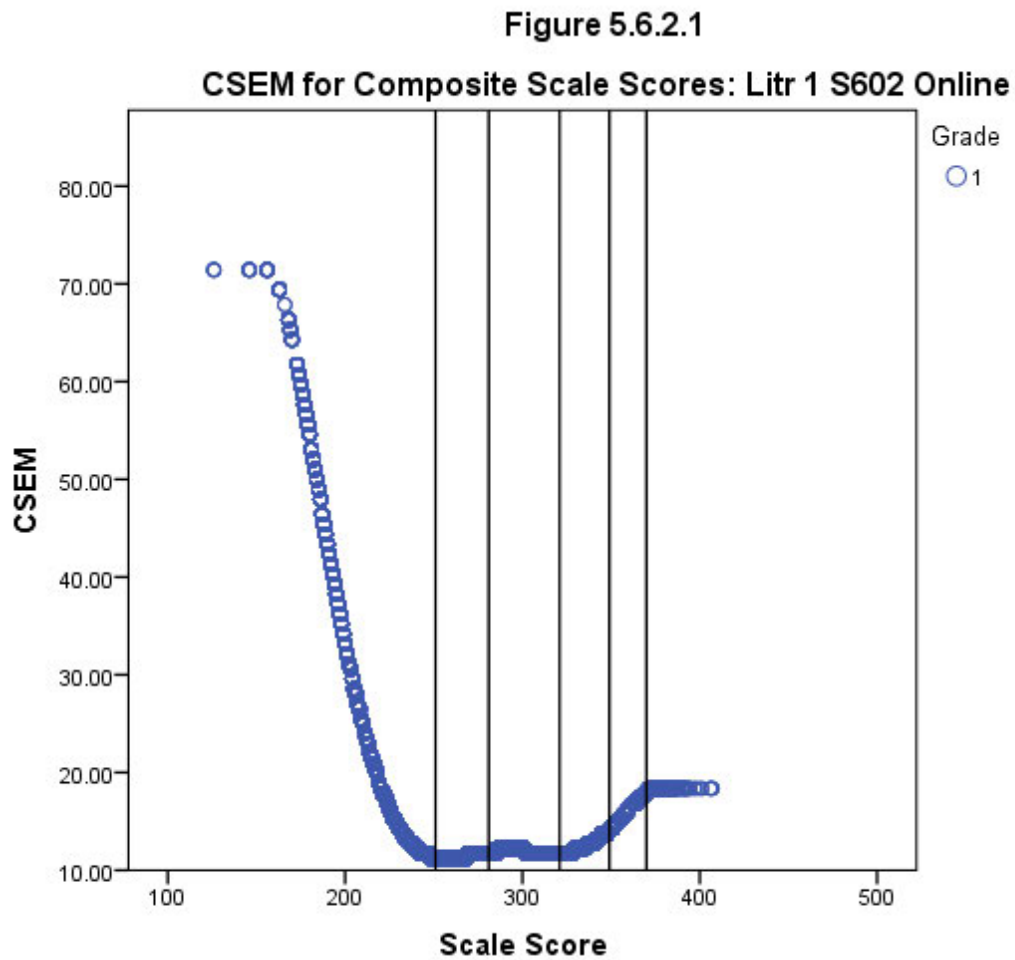


5.6.2 Literacy

5.6.2.1 Grade 1

Figure 5.6.2.1

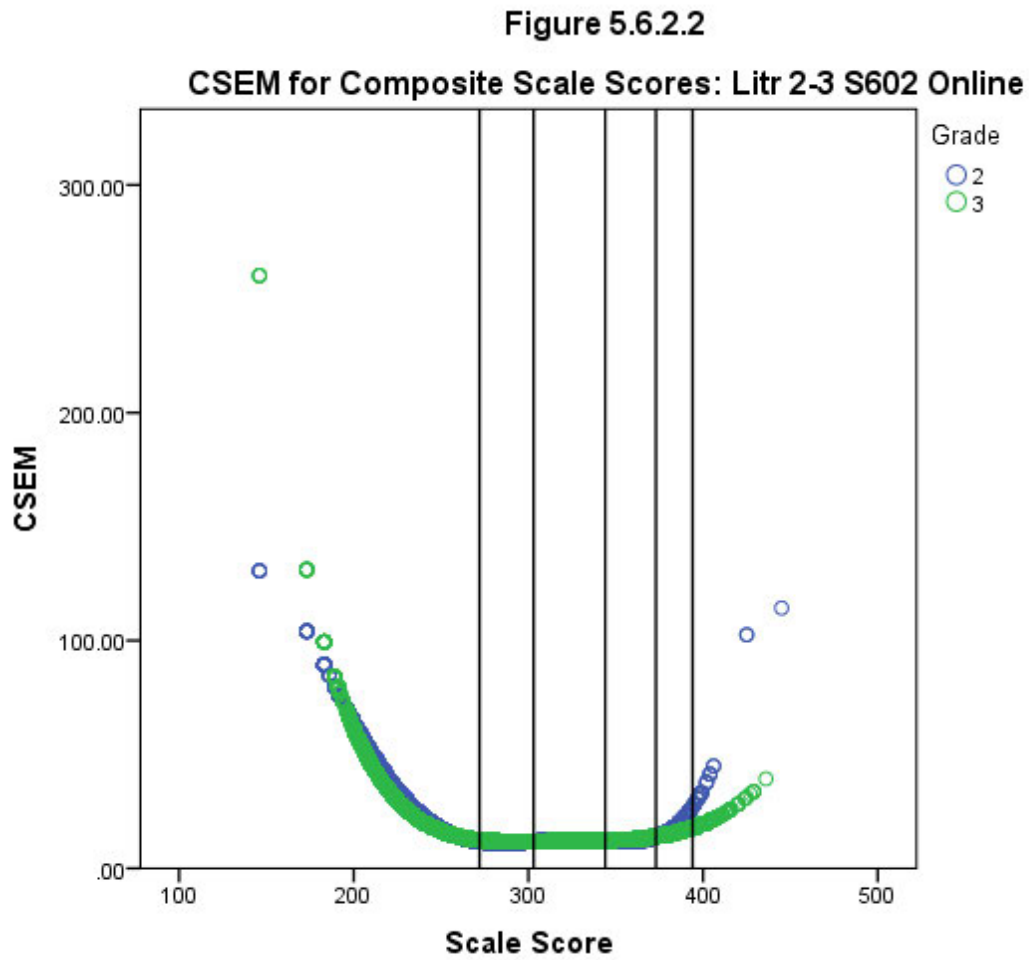
CSEM for Composite Scale Scores: Litr 1 S602 Online



5.6.2.2 Grades 2–3

Figure 5.6.2.2

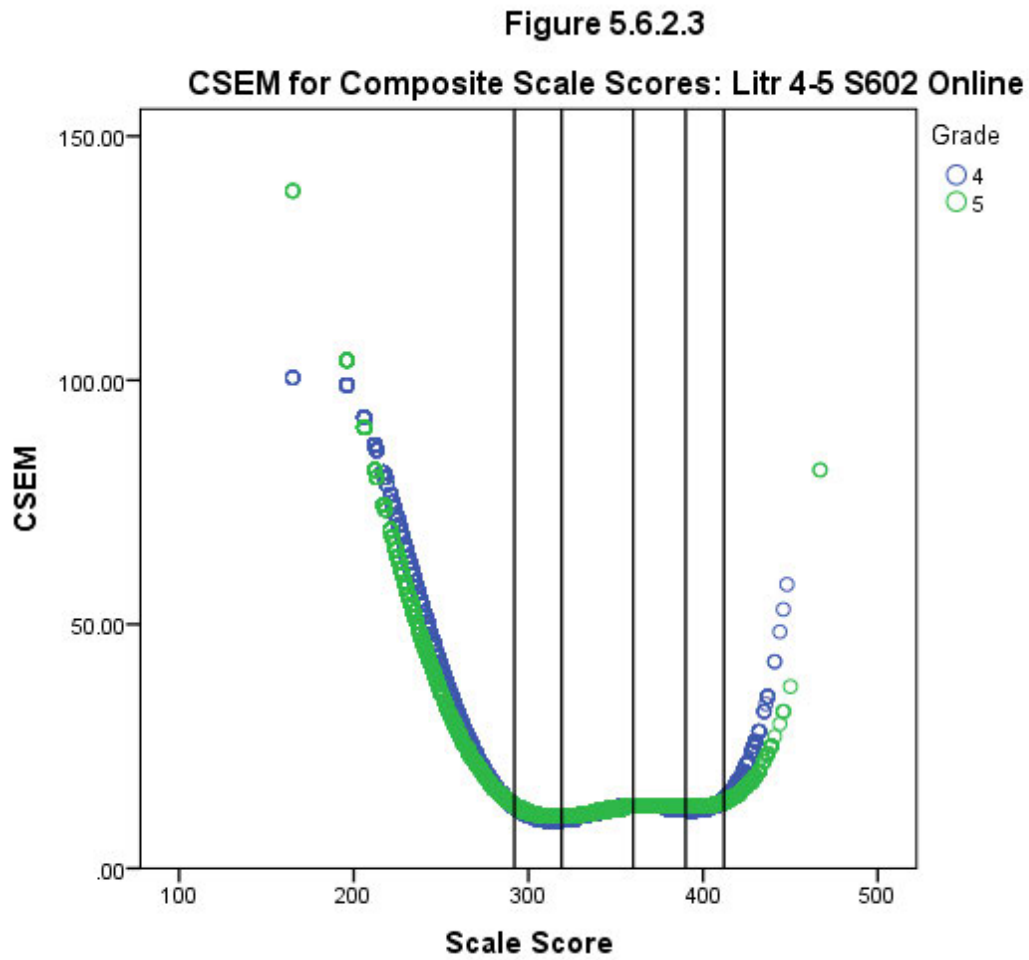
CSEM for Composite Scale Scores: Littr 2–3 S602 Online



5.6.2.3 Grades 4–5

Figure 5.6.2.3

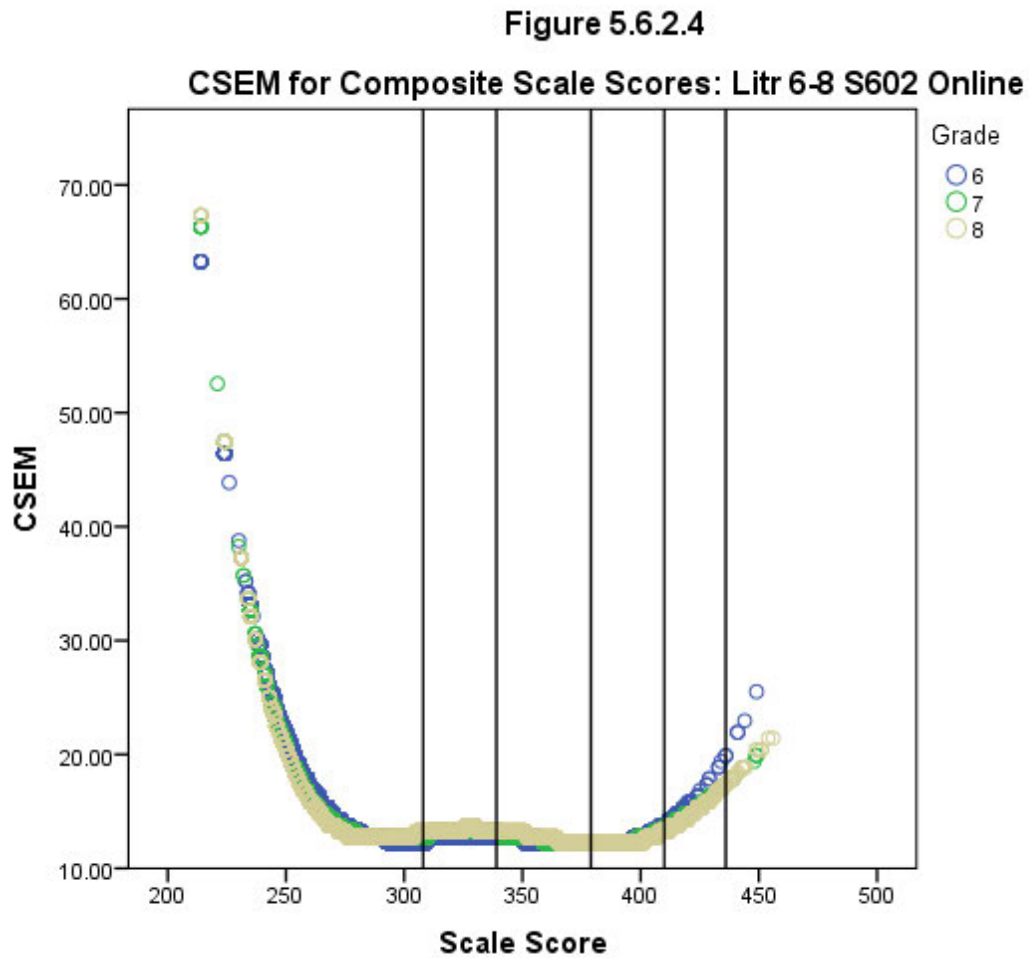
CSEM for Composite Scale Scores: Littr 4–5 S602 Online



5.6.2.4 Grades 6–8

Figure 5.6.2.4

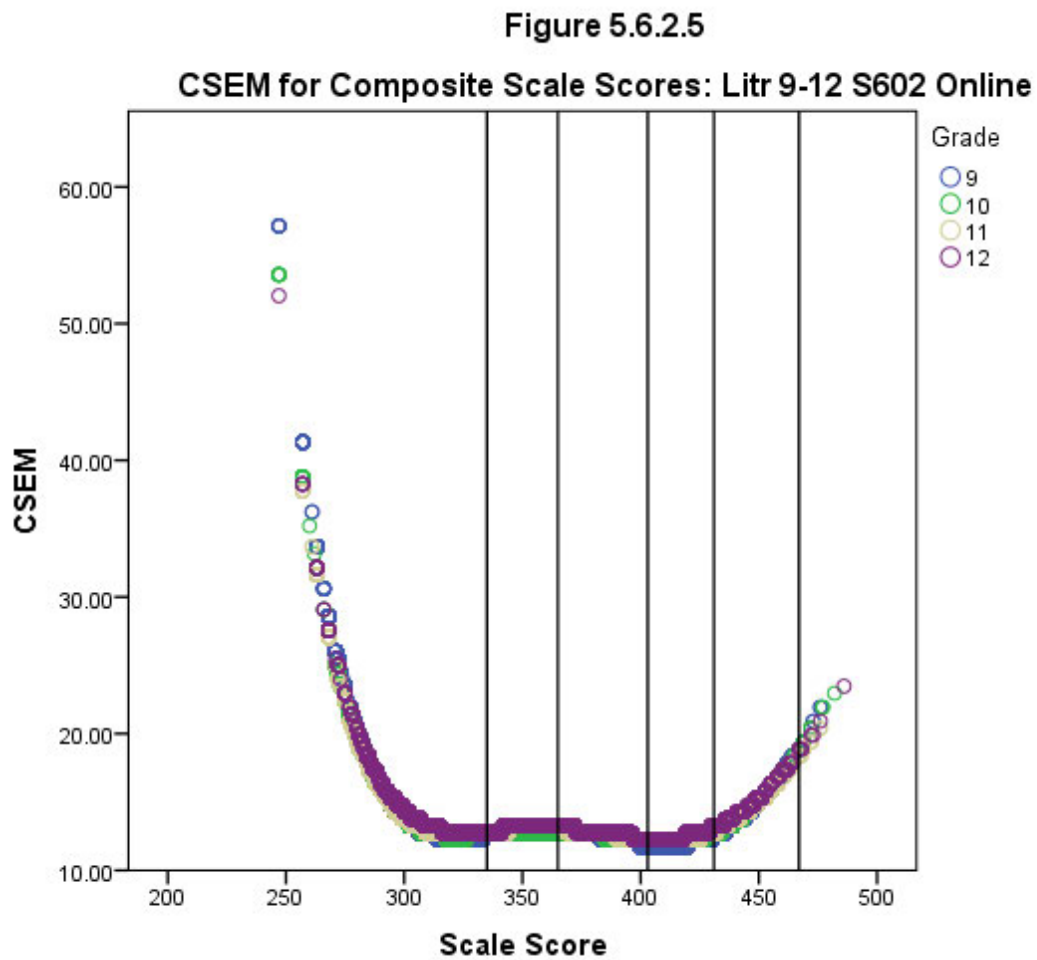
CSEM for Composite Scale Scores: Littr 6–8 S602 Online



5.6.2.5 Grades 9–12

Figure 5.6.2.5

CSEM for Composite Scale Scores: Litr 9–12 S602 Online

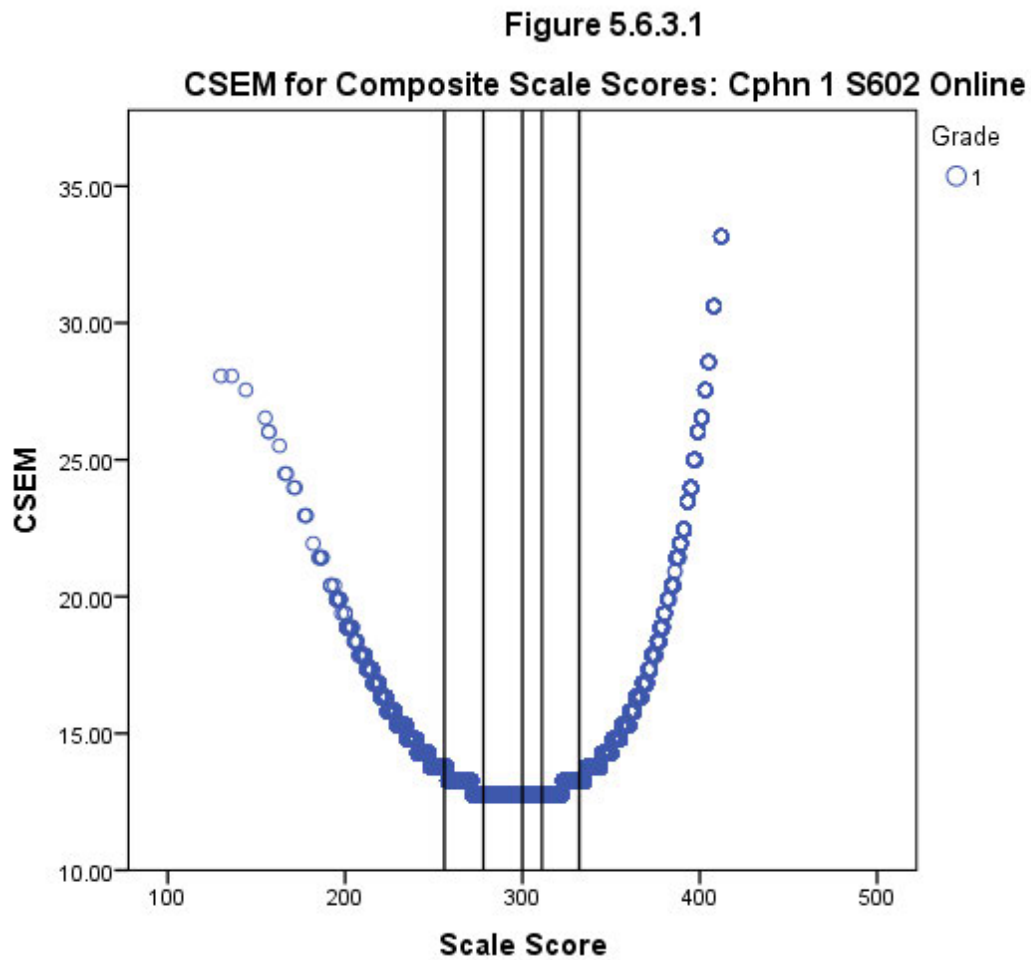


5.6.3 Comprehension

5.6.3.1 Grade 1

Figure 5.6.3.1

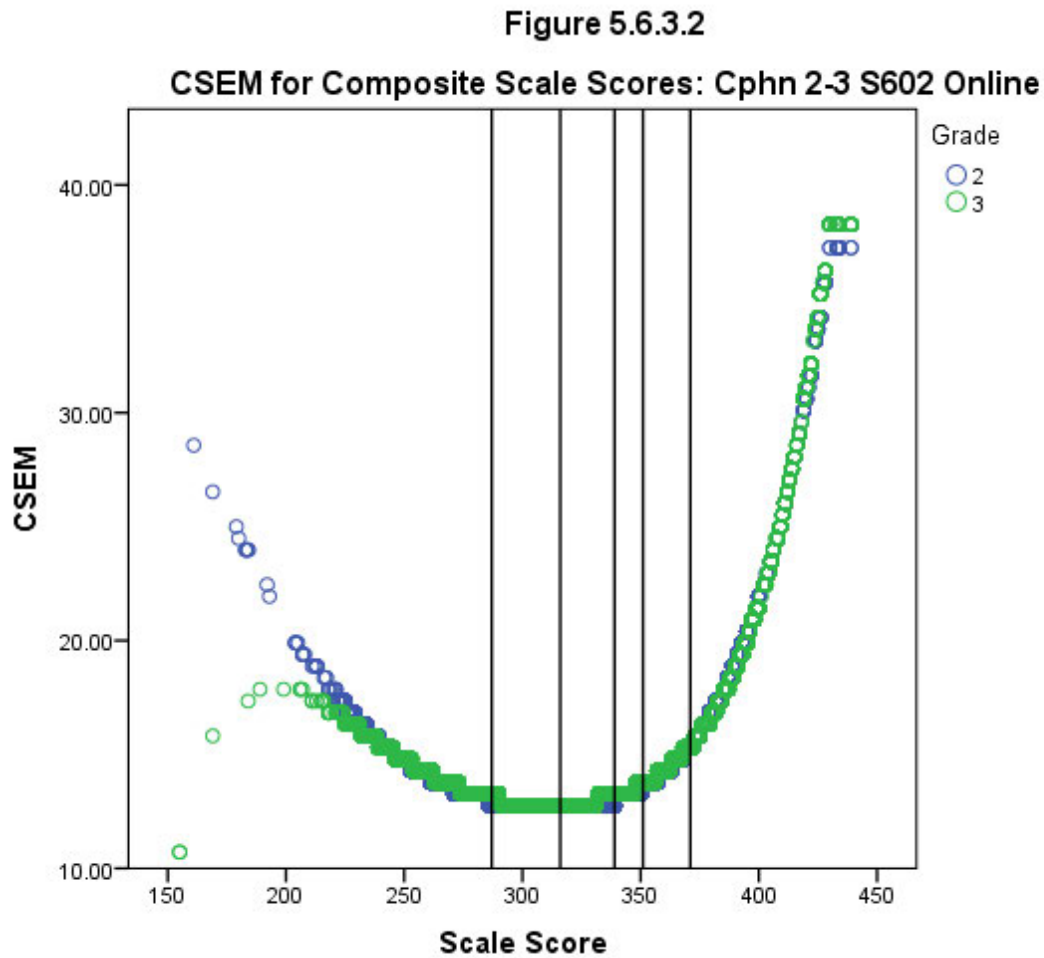
CSEM for Composite Scale Scores: Cphn 1 S602 Online



5.6.3.2 Grades 2–3

Figure 5.6.3.2

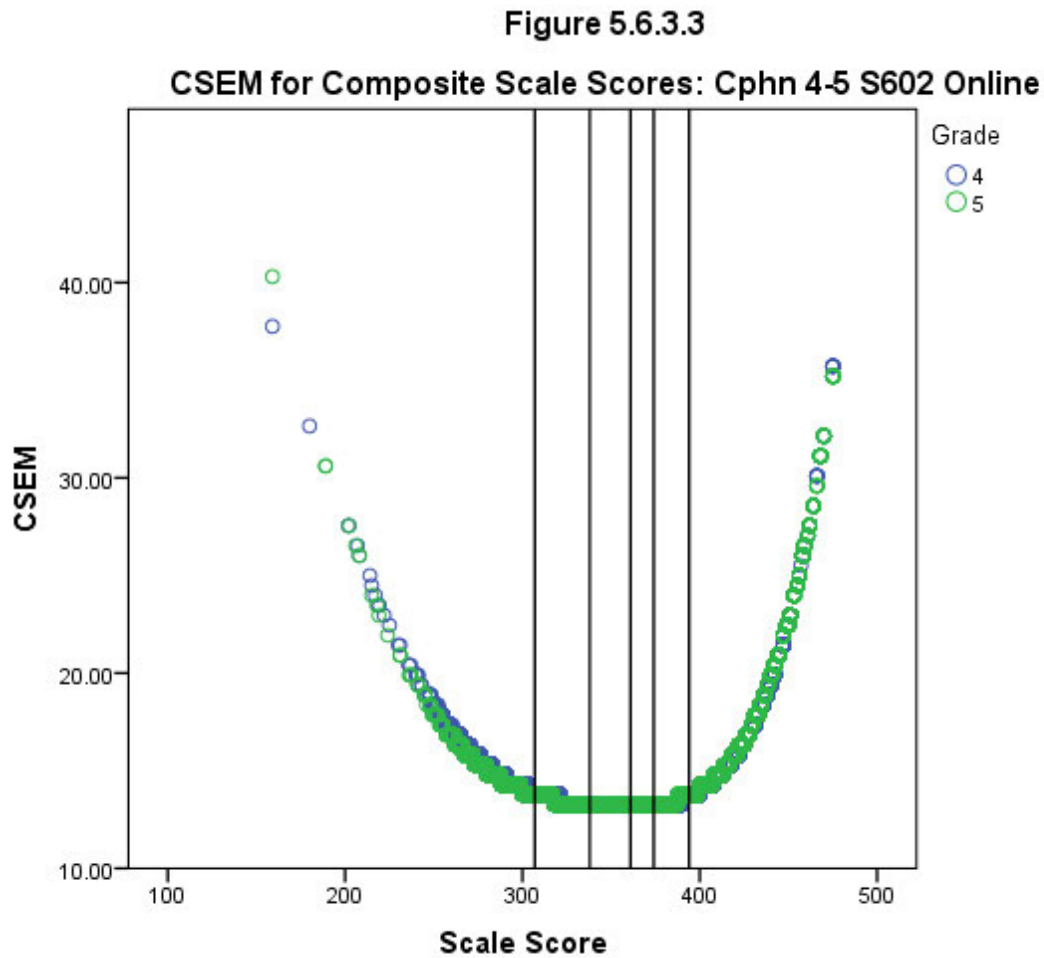
CSEM for Composite Scale Scores: Cphn 2–3 S602 Online



5.6.3.3 Grades 4–5

Figure 5.6.3.3

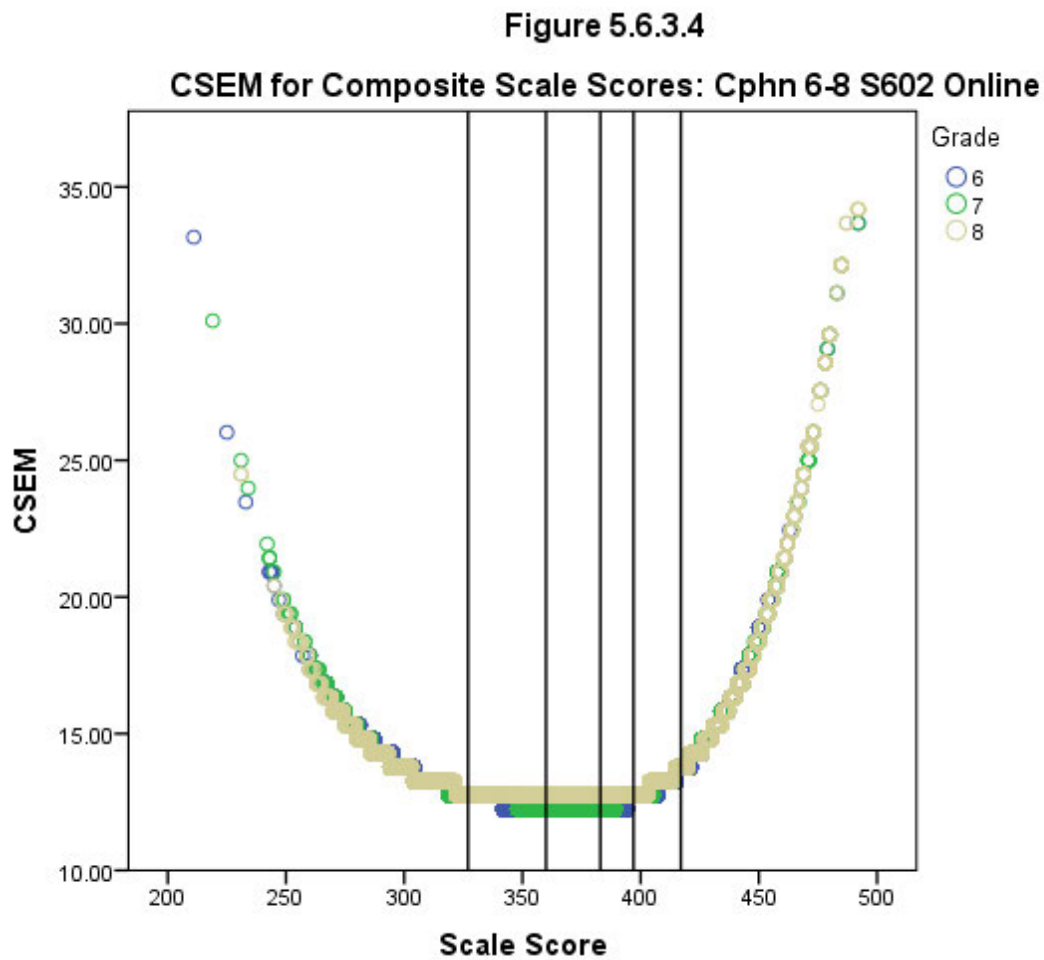
CSEM for Composite Scale Scores: Cphn 4–5 S602 Online



5.6.3.4 Grades 6–8

Figure 5.6.3.4

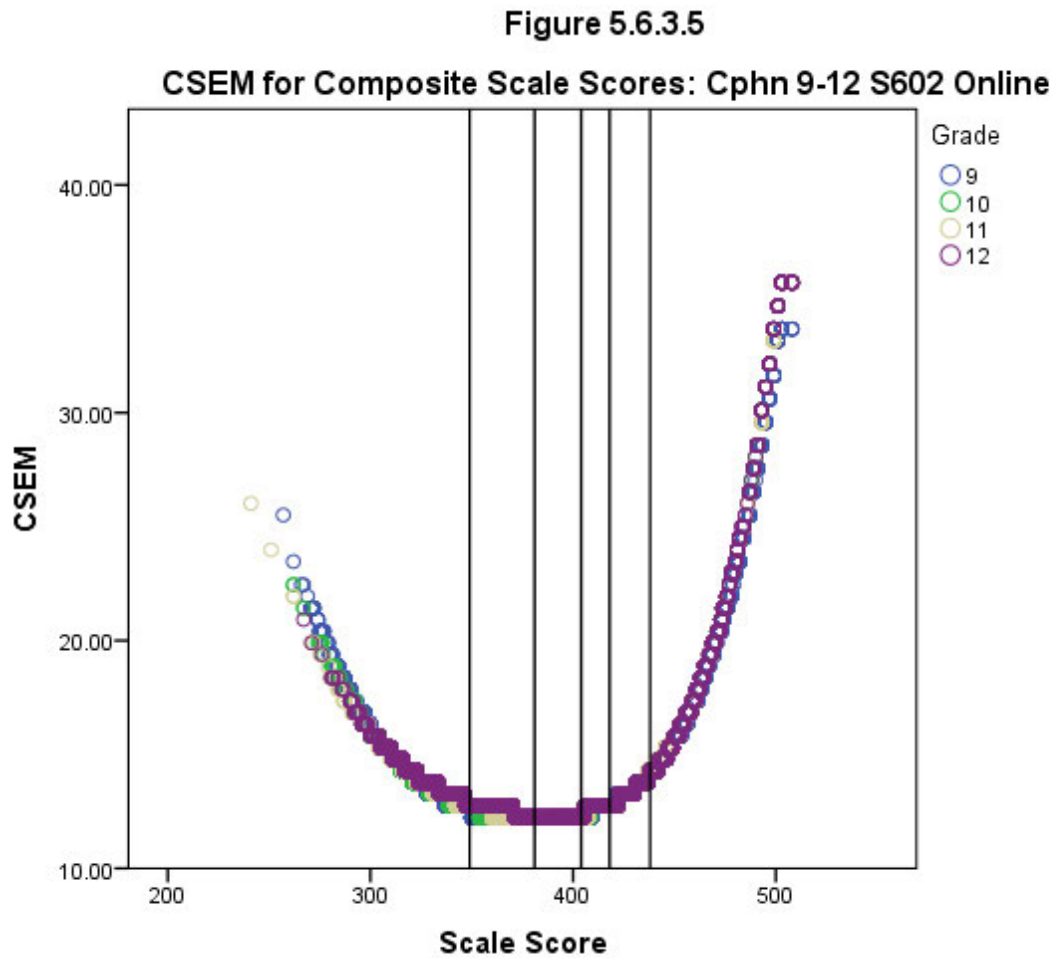
CSEM for Composite Scale Scores: Cphn 6–8 S602 Online



5.6.3.5 Grades 9–12

Figure 5.6.3.5

CSEM for Composite Scale Scores: Cphn 9–12 S602 Online

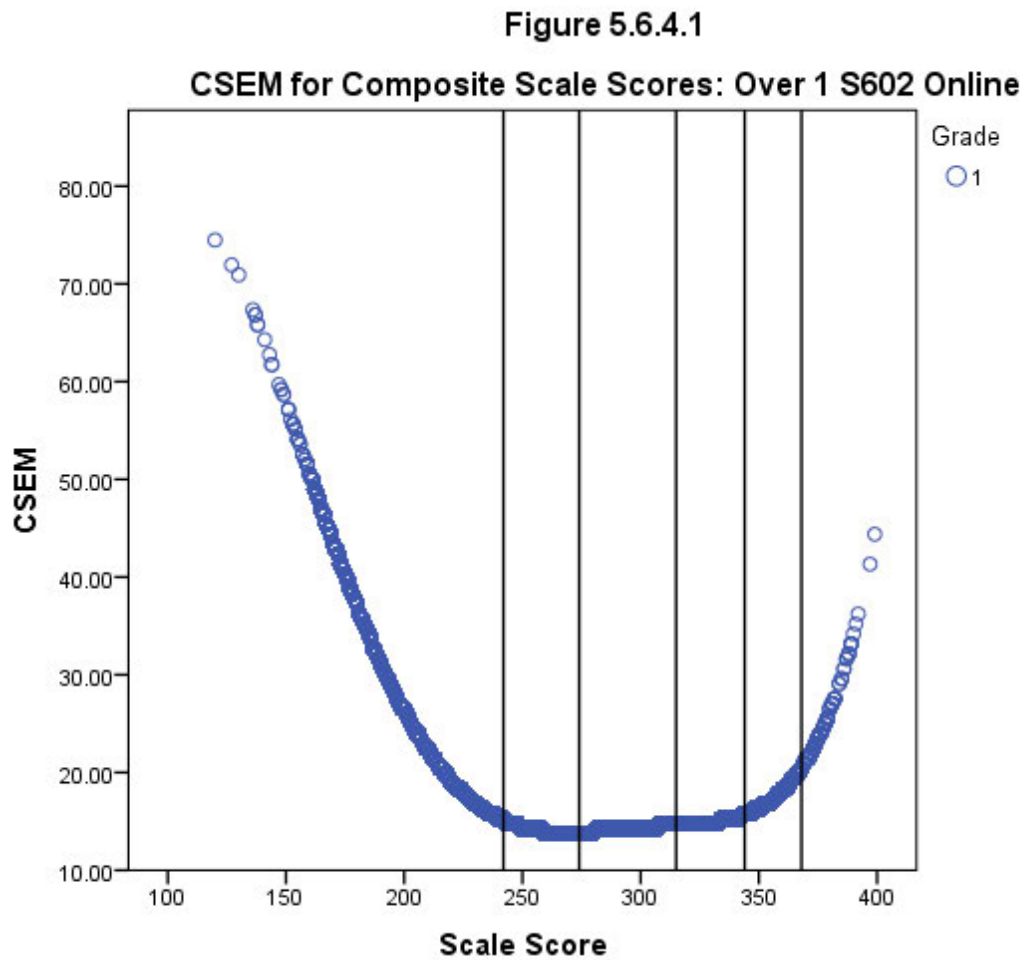


5.6.4 Overall

5.6.4.1 Grade 1

Figure 5.6.4.1

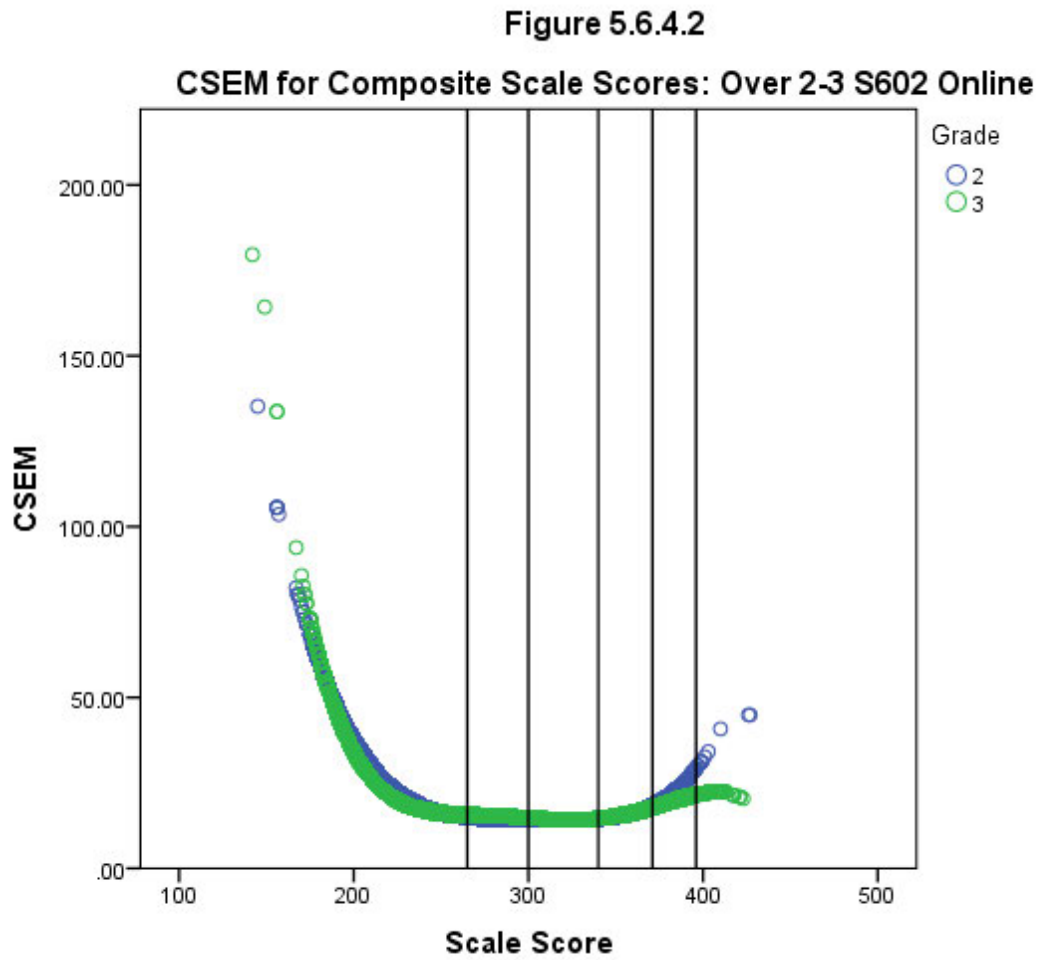
CSEM for Composite Scale Scores: Over 1 S602 Online



5.6.4.2 Grades 2–3

Figure 5.6.4.2

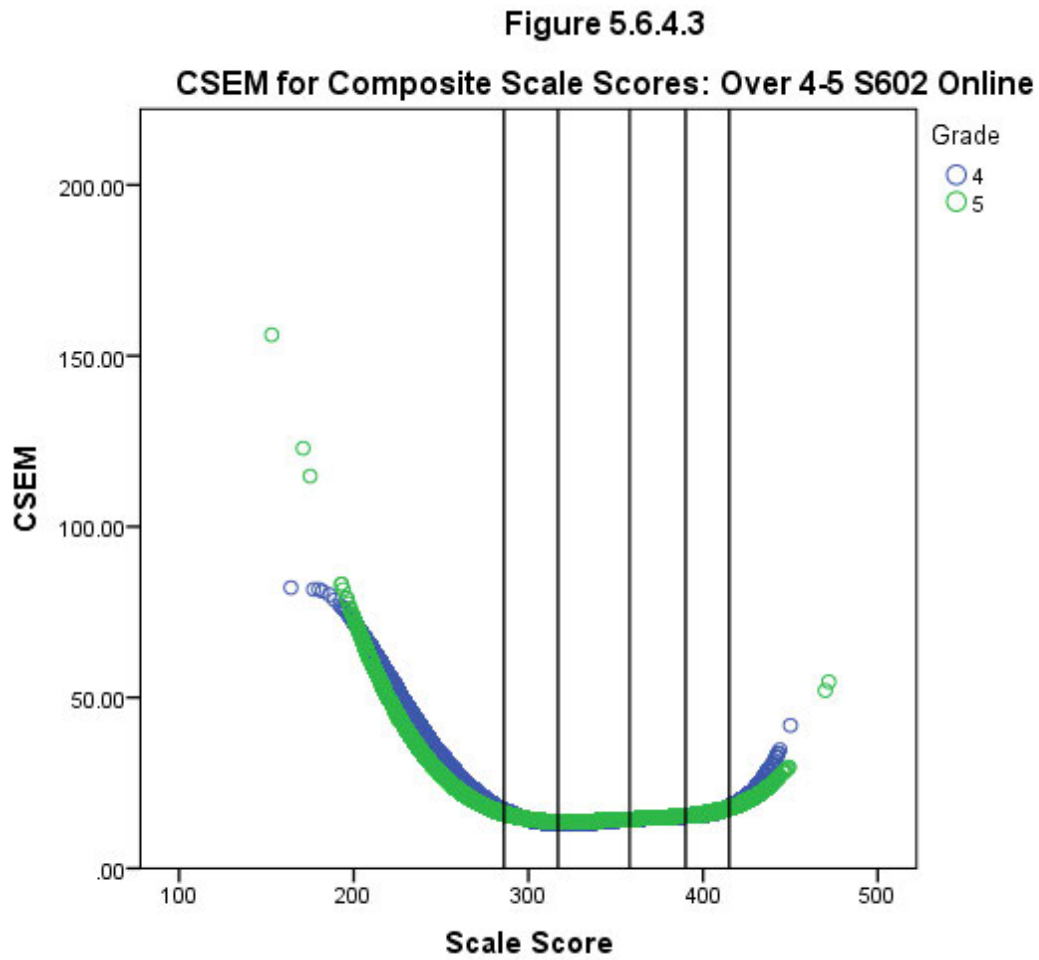
CSEM for Composite Scale Scores: Over 2–3 S602 Online



5.6.4.3 Grades 4–5

Figure 5.6.4.3

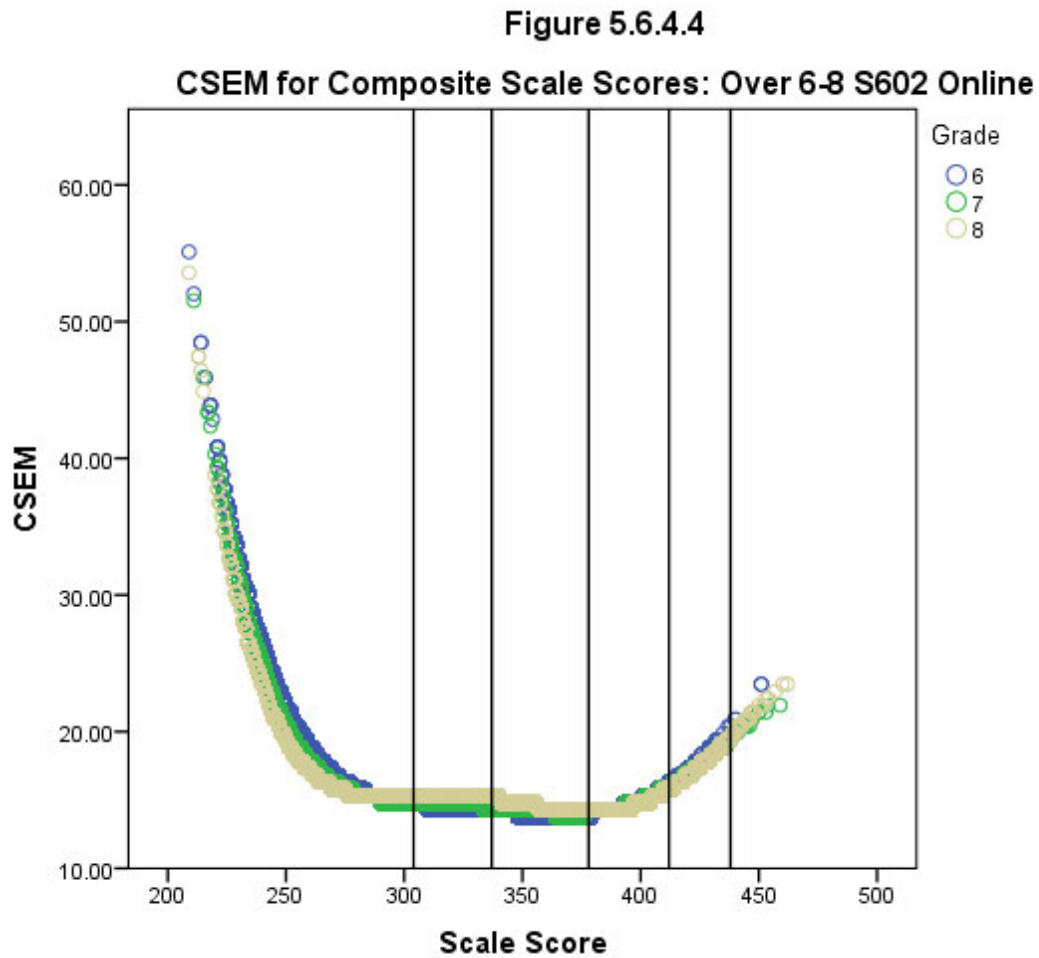
CSEM for Composite Scale Scores: Over 4–5 S602 Online



5.6.4.4 Grades 6–8

Figure 5.6.4.4

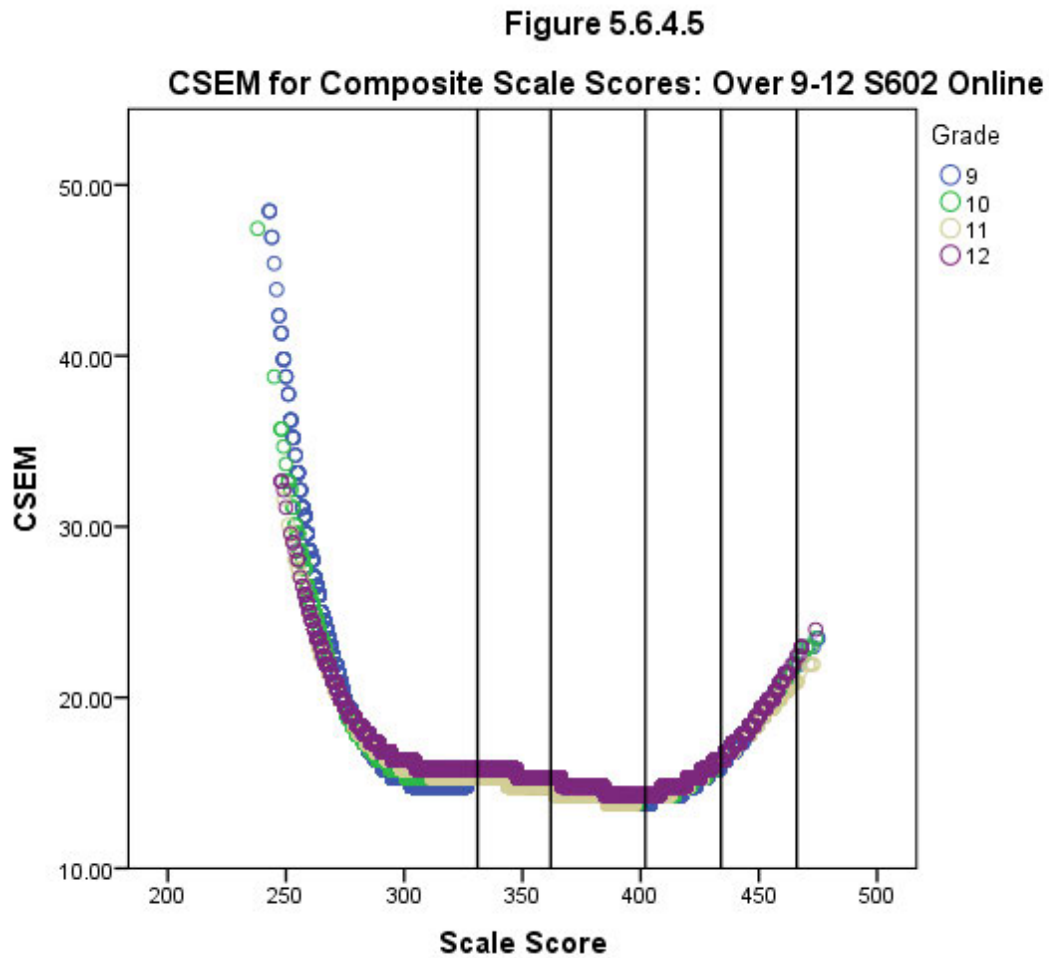
CSEM for Composite Scale Scores: Over 6–8 S602 Online



5.6.4.5 Grades 9–12

Figure 5.6.4.5

CSEM for Composite Scale Scores: Over 9–12 S602 Online



5.7 Accuracy and Consistency of Composites

One of the main purposes of the WIDA ACCESS program is to identify students' English language proficiency level concerning the WIDA ELD Standards. Because of the emphasis on classifying student performance, a question of interest is how accurately and consistently the ACCESS composite scale scores can classify students into WIDA proficiency categories determined by the 2016 ACCESS standard-setting process (Cook & MacGregor, 2017). Although states in the WIDA Consortium take into consideration one or more of the domain and composite scale scores when making accountability decisions, all WIDA Consortium states use the Overall composite scale score as the primary score when making classification decisions about students. Therefore, it is especially important to examine the accuracy and consistency of the classifications based on the **Overall composite scale scores** to help test users and policymakers judge the utility of this information and make decisions about score reporting (American Educational Research Association et al., 2014). The analyses utilize the methods that Livingston and Lewis (1995) and Young and Yoon (1998) outlined, as implemented in the software program BB-CLASS (Brennan, 2004; cf. also Lee et al., 2002).

The method and descriptions of the classification accuracy and consistency indices reported in this section appear in detail in Section 5.4. The only substantive methodological difference between the estimation of the classification accuracy and consistency of the domain scale scores versus the composite scale scores is that to estimate the classification accuracy and consistency of the composite scale scores, we first estimate the reliability of the composite scale scores using a stratified Cronbach's coefficient alpha, as described in Section 5.4.

For each composite, we present three tables. The first table reports the overall accuracy and the overall consistency indices for each grade. The second table reports the marginal classification accuracy indices based on the composite scale scores at the cut points for each grade. The third table reports the marginal classification consistency indices based on the composite scale scores at the cut points for each grade.

If we could not estimate the overall and marginal classification accuracy and consistency indices because there were fewer than 200 students in the proficiency level, we collapsed the affected proficiency level with the level below it and placed 'N/A' in the table for the affected proficiency level.

As noted in Section 5.4, assessment experts have issued very little guidance to aid in making judgments about the ideal or expected levels of decision consistency and accuracy needed for educational assessments. To help test users and policymakers interpret the results from our analyses, we report for each composite the range of these indices, highlighting the grade with the lowest classification accuracy and consistency indices for that composite. Since overall accuracy and consistency indices are summaries of the degree of classification accuracy and consistency for the composite scale scores across all proficiency level cut points, we also examine the marginal classification accuracy and consistency indices for these grades to identify the specific source(s) of low classification accuracy and consistency.

For the Oral composite, as shown in Table 5.7.1.1, the overall classification accuracy indices ranged from 0.651 to 0.757, and the overall classification consistency indices ranged from

0.547 to 0.667 across grades. The lowest overall classification accuracy and consistency indices were found for students in grade 5.

For the Literacy composite, as shown in Table 5.7.2.1, the overall classification accuracy indices ranged from 0.682 to 0.778, and the overall classification consistency indices ranged from 0.571 to 0.693 across grades. Grade 5 had the lowest overall classification accuracy and consistency indices.

For the Comprehension composite, as shown in Table 5.7.3.1, the overall classification accuracy indices ranged from 0.648 to 0.721, and the overall classification consistency indices ranged from 0.539 to 0.623 across grades. Grade 1 had the lowest overall classification accuracy and consistency indices.

For the Overall composite, as shown in Table 5.7.4.1, the overall classification accuracy indices ranged from 0.737 to 0.824, and the overall classification consistency indices ranged from 0.643 to 0.753 across grades. Grade 5 had the lowest overall classification accuracy and consistency indices.

The results reveal that grade 5 had the lowest overall classification accuracy and consistency indices for the Oral, Literacy, and Overall composites, while grade 1 had the lowest overall classification accuracy and consistency indices for the Comprehension composite.

From an accountability perspective, the most important indices for test users and policymakers to examine are the marginal classification accuracy and consistency indices. We report for each composite the range of the marginal classification accuracy and consistency indices for the composite scale scores across grades and then highlight the grade (and the cut point within that grade) that had the lowest marginal classification accuracy and the lowest consistency indices.

For the Oral composite, the marginal classification accuracy indices based on the scale scores at the cut points ranged from 0.892 to 0.998 (Table 5.7.1.2), and the marginal classification consistency indices ranged from 0.847 to 0.998 (Table 5.7.1.3). Grade 5, at the PL 4/5 cut point, had the lowest marginal classification accuracy and consistency indices. Note that grade 5 also had the lowest overall classification accuracy and consistency indices for the Oral composite. The low marginal classification accuracy and consistency at the PL 4/5 cut point appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency indices for the grade 5 Oral composite are still in the high 0.80 to mid-0.90 range.

For the Literacy composite, the marginal classification accuracy indices based on the scale scores at the cut points ranged from 0.873 to 0.999 (Table 5.7.2.2), and the marginal classification consistency indices ranged from 0.822 to 0.999 (Table 5.7.2.3). Grade 5, at the PL 3/4 cut point, had the lowest marginal classification accuracy and consistency indices. Note that grade 5 also had the lowest overall classification accuracy and consistency indices for the Literacy composite. The low marginal classification accuracy and consistency at the PL 3/4 cut point appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal accuracy and consistency indices for the grade 5 Literacy composite are still in the high 0.80 to mid-0.90 range.

For the Comprehension composite, the marginal classification accuracy indices based on the scale scores at the cut points ranged from 0.900 to 0.975 (Table 5.7.3.2), and the marginal classification consistency indices ranged from 0.859 to 0.963 (Table 5.7.3.3). Grade 1, at the PL 2/3 cut point, had the lowest marginal classification accuracy and consistency indices. Note that grade 1 also had the lowest overall classification accuracy and consistency indices for the Comprehension composite. The low marginal classification accuracy and consistency at the PL 2/3 cut point appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal accuracy and consistency indices for the grade 1 Comprehension composite are still in the high 0.80 to mid-0.90 range.

For the Overall composite, the marginal classification accuracy indices based on the scale scores at the cut points ranged from 0.912 to 0.999 (Table 5.7.4.2), and the marginal classification consistency indices ranged from 0.876 to 0.999 (Table 5.7.4.3). Grade 5 had the lowest marginal classification accuracy at the PL 3/4 cut point. Note that grade 5 also had the lowest overall classification accuracy and consistency indices for the Overall composite. The low marginal classification accuracy and consistency at the PL 3/4 cut points appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal accuracy and consistency indices for the grade 5 Overall composite are still in the high 0.80 to mid-0.90 range.

When we compared the overall and marginal classification accuracy and consistency indices for the composites for a particular grade, we saw that in many instances they told the same story (i.e., for a given grade, if the overall classification accuracy and consistency indices were low, then the marginal classification accuracy and consistency indices also tended to be low). This was especially true for grade 5 for three of the four composites (Oral, Literacy, and Overall). Grade 5 had the lowest overall and marginal classification accuracy and consistency indices for these composites. Similarly, grade 1 had the lowest overall and marginal classification accuracy and consistency indices for the Comprehension composite. In addition, the lowest marginal classification accuracy and consistency based on the composite scale scores occurred at the PL 2/PL 3, PL 3/PL 4, and PL 4/PL 5 cut points. A higher number of proficiency levels typically results in cut points that are closer to each other than if there were a smaller number of proficiency levels. We would expect marginal classification accuracy and consistency to vary for different ability levels due to variations in measurement accuracy. That is, the further away the students' composite scale scores are from the cut points, the smaller the classification errors would be, or the more accurate the classification decisions would be. With many proficiency levels, there are more student composite scale scores near the cut points than there would be if there were fewer with only two proficiency levels. Therefore, the higher the number of proficiency levels, the higher the probability that students would be misclassified (Ercikan & Julian, 2002). The marginal classification accuracy and consistency indices based on the composite scale scores for cut points that are in the middle range tend to be lower than for other cut points, as we might expect.

Assessment experts have issued little guidance to aid in making judgments about the ideal or expected levels of decision consistency and accuracy needed for educational assessments that report composite scale scores. From an accountability perspective, the most important indices are the marginal classification accuracy and consistency indices. The marginal classification

accuracy and consistency indices were at or above 0.822 for all four composites. Additionally, the marginal classification accuracy and consistency indices were at or above 0.876 for the Overall composite scale score, which is the primary score that WIDA Consortium states use when making accountability decisions.

5.7.1 Oral

Table 5.7.1.1

Overall Accuracy and Consistency of Classification Indices: Oral S602 Online

Grade	Accuracy	Consistency
1	0.714	0.613
2	0.722	0.624
3	0.691	0.594
4	0.670	0.558
5	0.651	0.547
6	0.736	0.634
7	0.720	0.617
8	0.705	0.604
9	0.757	0.663
10	0.754	0.662
11	0.749	0.657
12	0.757	0.667

Table 5.7.1.2

Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Oral S602 Online

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.947	0.923	0.910	0.941	0.990
2	0.952	0.914	0.910	0.948	0.994
3	0.955	0.919	0.895	0.921	0.992
4	0.975	0.954	0.916	0.898	0.924
5	0.969	0.951	0.910	0.892	0.926
6	0.961	0.928	0.905	0.949	0.991
7	0.955	0.928	0.906	0.939	0.991
8	0.953	0.929	0.906	0.924	0.989
9	0.942	0.917	0.918	0.980	0.998
10	0.940	0.916	0.920	0.978	0.998
11	0.937	0.913	0.921	0.976	0.997
12	0.938	0.912	0.925	0.983	N/A

Table 5.7.1.3

Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Oral S602 Online

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.925	0.891	0.874	0.915	0.989
2	0.932	0.879	0.873	0.932	0.994
3	0.936	0.885	0.852	0.905	0.991
4	0.966	0.933	0.882	0.857	0.897
5	0.957	0.928	0.874	0.847	0.909
6	0.945	0.898	0.867	0.926	0.990
7	0.937	0.898	0.868	0.913	0.988
8	0.933	0.899	0.868	0.898	0.986
9	0.918	0.882	0.884	0.973	0.998
10	0.916	0.881	0.887	0.971	0.998
11	0.911	0.878	0.888	0.970	0.997
12	0.911	0.875	0.892	0.978	N/A

5.7.2 Literacy

Table 5.7.2.1

Overall Accuracy and Consistency of Classification Indices: Litr S602 Online

Grade	Accuracy	Consistency
1	0.778	0.693
2	0.740	0.642
3	0.715	0.611
4	0.685	0.576
5	0.682	0.571
6	0.776	0.688
7	0.763	0.671
8	0.750	0.655
9	0.740	0.640
10	0.746	0.646
11	0.740	0.640
12	0.750	0.652

Table 5.7.2.2**Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Litr S602 Online**

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.896	0.913	0.976	0.994	0.999
2	0.943	0.895	0.911	0.989	0.999
3	0.947	0.907	0.885	0.974	0.998
4	0.950	0.919	0.878	0.935	0.991
5	0.950	0.921	0.873	0.934	0.993
6	0.934	0.902	0.943	0.996	N/A
7	0.935	0.903	0.932	0.993	N/A
8	0.934	0.908	0.918	0.990	N/A
9	0.945	0.903	0.916	0.978	0.998
10	0.948	0.901	0.918	0.978	N/A
11	0.941	0.898	0.923	0.977	N/A
12	0.933	0.894	0.936	0.986	N/A

Table 5.7.2.3**Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Litr S602 Online**

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.854	0.876	0.967	0.992	0.999
2	0.918	0.854	0.876	0.985	0.999
3	0.925	0.869	0.838	0.962	0.998
4	0.928	0.886	0.829	0.910	0.988
5	0.929	0.889	0.822	0.905	0.990
6	0.908	0.862	0.918	0.995	N/A
7	0.909	0.863	0.904	0.990	N/A
8	0.907	0.869	0.885	0.985	N/A
9	0.923	0.863	0.881	0.967	0.998
10	0.926	0.861	0.885	0.968	N/A
11	0.917	0.857	0.890	0.969	N/A
12	0.905	0.851	0.909	0.981	N/A

5.7.3 Comprehension

Table 5.7.3.1

Overall Accuracy and Consistency of Classification Indices: Cphn S602 Online

Grade	Accuracy	Consistency
1	0.648	0.539
2	0.700	0.596
3	0.677	0.576
4	0.699	0.603
5	0.679	0.582
6	0.721	0.623
7	0.701	0.601
8	0.685	0.585
9	0.705	0.606
10	0.699	0.600
11	0.697	0.598
12	0.699	0.600

Table 5.7.3.2

Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Cphn S602 Online

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.921	0.900	0.914	0.932	0.966
2	0.960	0.916	0.916	0.935	0.967
3	0.941	0.919	0.919	0.927	0.955
4	0.973	0.943	0.928	0.916	0.927
5	0.961	0.943	0.922	0.910	0.929
6	0.955	0.920	0.925	0.941	0.975
7	0.947	0.925	0.924	0.934	0.965
8	0.947	0.928	0.920	0.925	0.956
9	0.957	0.922	0.921	0.934	0.964
10	0.954	0.922	0.921	0.933	0.961
11	0.944	0.921	0.924	0.936	0.962
12	0.941	0.920	0.924	0.940	0.966

Table 5.7.3.3

Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Cphn S602 Online

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.889	0.859	0.880	0.904	0.950
2	0.943	0.882	0.882	0.908	0.953
3	0.918	0.886	0.886	0.898	0.936
4	0.962	0.920	0.897	0.884	0.897
5	0.946	0.919	0.890	0.877	0.899
6	0.936	0.888	0.895	0.917	0.963
7	0.926	0.894	0.893	0.908	0.949
8	0.926	0.898	0.888	0.896	0.936
9	0.940	0.890	0.889	0.908	0.949
10	0.936	0.890	0.889	0.907	0.944
11	0.922	0.889	0.893	0.910	0.945
12	0.917	0.887	0.894	0.915	0.951

5.7.4 Overall

Table 5.7.4.1

Overall Accuracy and Consistency of Classification Indices: Over S602 Online

Grade	Accuracy	Consistency
1	0.817	0.744
2	0.801	0.723
3	0.779	0.694
4	0.744	0.649
5	0.737	0.643
6	0.824	0.753
7	0.812	0.737
8	0.803	0.725
9	0.808	0.732
10	0.811	0.735
11	0.807	0.730
12	0.811	0.737

Table 5.7.4.2**Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Over S602 Online**

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.934	0.920	0.971	0.992	0.999
2	0.959	0.925	0.932	0.985	N/A
3	0.962	0.932	0.913	0.973	N/A
4	0.971	0.949	0.915	0.922	0.985
5	0.969	0.950	0.912	0.913	0.991
6	0.962	0.932	0.937	0.993	N/A
7	0.959	0.933	0.932	0.988	N/A
8	0.957	0.935	0.927	0.985	N/A
9	0.958	0.931	0.937	0.983	N/A
10	0.957	0.929	0.940	0.986	N/A
11	0.954	0.927	0.942	0.984	N/A
12	0.950	0.923	0.950	0.988	N/A

Table 5.7.4.3**Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Over S602 Online**

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
1	0.907	0.888	0.959	0.991	0.999
2	0.942	0.894	0.904	0.983	N/A
3	0.946	0.903	0.877	0.966	N/A
4	0.959	0.928	0.880	0.890	0.983
5	0.956	0.928	0.876	0.883	0.989
6	0.946	0.904	0.911	0.992	N/A
7	0.942	0.906	0.904	0.985	N/A
8	0.939	0.907	0.897	0.980	N/A
9	0.941	0.902	0.911	0.978	N/A
10	0.940	0.900	0.914	0.981	N/A
11	0.935	0.897	0.917	0.981	N/A
12	0.929	0.892	0.928	0.988	N/A

6. Quality Control

6.1 Content Development Quality Control

The Center for Applied Linguistics (CAL) utilizes educators and other consultants at a number of phases throughout the test development cycle. These educators and consultants are recruited, vetted, and trained by CAL and/or WIDA and make crucial contributions to these phases of the test development cycle. The phases of development in which educators or consultants are involved, as well as the procedures and criteria for recruitment and training, are described below.

Theme Generation: During theme generation, CAL and WIDA recruit educators to generate raw ideas to be used in new item development. Educators with ESL or content-area expertise and two or more years of teaching experience in a WIDA state (in the grade-level cluster for which they will generate themes) are invited to participate. Recruitment also focuses on a geographical distribution of educators from across the consortium. Upon selection, educators participate in a short training that introduces the theme generation process, along with how to understand the item specifications that they use to generate themes.

Item Writing: CAL recruits professional item writers to generate raw item/task content based on the ideas from theme generation. To recruit item writers, CAL has a standing announcement on its website asking prospective item writers to submit their resume and fill out a survey describing their past item-writing experience. CAL selects individuals with significant experience in writing items, both in large-scale assessment programs (ESL/EFL or ELA) and in other contexts (e.g., writing items for assessment programs in university-based ESL programs).

Item writers undergo a 90-minute orientation prior to beginning item writing. This training focuses on the item specifications, the process and procedures, the item writing checklist, the acceptance criteria for the items, and the security protocols. Item writers also receive an item writing handbook, which formalizes the content of the orientation, along with assignment of themes to develop and the associated item specifications. After the orientation, CAL language testing specialists and managers provide feedback to the item writers on the items, focusing on alignment with the item writing checklist and the item specifications. After completion of item writing for a given development cycle, item writers are evaluated by CAL staff for their compliance with the requirements and the quality of their items.

Standards Expert Review: After items have been drafted by item writers, CAL language testing specialists review all of the raw content internally. This review focuses on determining which sets of items will move on to further development and which will be discontinued, based on criteria from an item review checklist. The language testing specialists then do minor editing and formatting to the items to make sure that they are complete, with no stray comments or other editorial notes from previous drafts, and they produce a short questionnaire for each set of items that becomes part of Standards Expert review. The purpose of Standards Expert review is to ensure that the items are appropriate for the grade level and intended difficulty level in terms of both the content and the language, and the items have not drifted from their

intended target between theme generation and item writing. The questionnaires produced by CAL's language testing specialists guide the Standards Experts through the review process, asking questions specific to the purpose of this review.

Educators are recruited jointly by CAL and WIDA to serve as Standards Experts; educators with ESL or content-area expertise and two or more years of teaching experience in a WIDA state are invited to participate. Recruitment also focuses on a geographical distribution of educators from across the consortium. Standards Experts receive written instructions and a questionnaire to complete for each set of items they review.

Bias & Sensitivity and Content Review: After Standards Expert review has been completed, all items undergo an additional phase of review and revision internal to CAL, leading up to Bias & Sensitivity and Content Review. These are technically two separate reviews, although a single recruitment effort is conducted by WIDA, and the reviews occur consecutively in a single week (generally 3 days for Content review followed by 2 days for Bias & Sensitivity review). As with other reviews, educators for Content review must have at least 2 years of ESL teaching experience (with a preference for content-area experience as well). Recruitment also focuses on selecting educators with a variety of cultural and linguistic backgrounds and obtaining a geographical distribution of educators from across the consortium. Recruitment for Bias & Sensitivity review focuses on selecting educators with culturally and linguistically diverse backgrounds who have experience interacting with English learners from a range of cultural, regional, religious, linguistic, ethnic, and socioeconomic backgrounds.

At the beginning of both Bias & Sensitivity and Content review meetings, CAL and WIDA staff conduct an intensive training to orient the reviewers to the specific purpose of the review (Bias & Sensitivity or Content), how to use the review checklist and what to look for in the review, and the procedures and security protocols for the review. Then, the reviews are conducted in breakout groups by grade-level cluster (or combinations of grade-level clusters; for example, Bias & Sensitivity review of grade 1 and grades 2–3 is often combined). Although Bias & Sensitivity and Content reviews are generally held in-person, the reviews for the Writing domain occur virtually each year due to timeline constraints. For both the in-person and virtual contexts, CAL and WIDA facilitators are present in each breakout group to guide the educators in their reviews of the materials.

Writing Tryouts: For the Writing domain, all tasks in the Writing test are subject to tryouts in the field. The Writing tryouts only occur once the tasks have been through a thorough Bias & Sensitivity and Content review and subsequent revision. CAL and WIDA recruit educators who are willing to administer the Writing tasks to their students; these educators are classroom ESL or content teachers who work with ELs. All students who participate are required to have parent/guardian consent.

Once the students complete the Writing tasks, both the students and educators fill out questionnaires. Student questionnaires focus on whether the students understood the task, their engagement with the task, and their ability to complete the task; educator surveys ask the teachers to evaluate the effectiveness of the task input, the appropriateness of the task, the comparability of the task with other classroom-based writing tasks, and the ability of the students to complete the task.

CAL provides the teachers with a number of documents outlining the procedures for administering the tasks, recording student responses to the tasks, recording student and teacher responses to the questionnaires, and protecting the personally identifiable information of the students. CAL staff are also available throughout the tryouts process to answer any questions the teachers might have. Following the Writing tryouts, CAL specialists review the writing responses both qualitatively and quantitatively, providing WIDA with a report on how the Writing tasks performed.

6.2 *Test Administration Quality Control*

This section describes how WIDA monitors test administration to ensure standardized test administration procedures are implemented with fidelity across districts and schools. To support standardized administrations, WIDA provides test administrators with a series of resources, such as a test administration manual, a training course, and a Test Administration Script for each assessment.

Qualifications of Test Administrators: Before, during, and after a state’s testing window, educators hold various roles to ensure all tasks are carried out for successful test administration. These roles include test coordinators at the district and school level and test administrators. The test administrator administers and monitors the test, and is also responsible for managing student data prior to, during, and after testing.

WIDA has worked directly with each state education agency to develop the ACCESS for ELLs Checklist for the school year. This list highlights all tasks that need to be completed before, during, and after testing within a school or district and outlines which tasks are assigned to Test Coordinators at the district and school level and to Test Administrators. It also provides additional guidance that a state expects test administrators to follow as they prepare for and administer the ACCESS for ELLs suite of assessments.

Test administrators are responsible for reviewing each state’s checklist in detail prior to completing any training and for working with the district or school test coordinator to complete these tasks. The state’s checklist can be found in the training course and on each state’s WIDA webpage.

The training course within the WIDA Secure Portal is where educators can access both training to become certified to administer ACCESS for ELLs as well as additional materials and resources to assist administrators and coordinators before, during, and after each state’s testing window. WIDA user accounts provide access to the training course and Facilitator Toolkit within the WIDA Secure Portal. Educators must pass an administration quiz at the end of the training with a score of 80% or higher. WIDA recommends taking the quiz immediately after completing the training. There is no limit to the number of times educators can attempt the quiz. Once individuals pass an administration quiz, training certificates within the WIDA Secure Portal are updated to reflect their status as a certified test administrator for that component of the assessment suite.

Paper Testing (for Writing Grades 1–3): Depending on state, district, and school policy, not all test administrators will be responsible for initially labeling and/or bubbling booklets.

However, it is the responsibility of all test administrators and test coordinators to ensure that correct and complete information is either labeled or bubbled in each student booklet. Each state's ACCESS for ELLs Checklist has more information on who is responsible for each task related to materials management in the state.

To ensure all booklets have the detailed and necessary information needed to score, all test administrators must adhere to the following:

- Prior to administration
 - Review labels and/or bubbled information to ensure all student information is accurate.
 - Complete labeling or bubbling if needed.
- During administration
 - Distribute the test booklets, as applicable, to the correct students.
 - Verify that students have been given their assigned booklet.
- Immediately following administration
 - Collect all material from all students.
 - Review student test booklets once more for any errors or discrepancies in student information.
 - Confirm all necessary fields are completed and all necessary labels are correctly adhered to student test booklets.
 - Ensure all booklets are in proper condition to be returned, with no loose or damaged pages.
 - Return test materials to a test coordinator or store the booklets in a secure area until they can be handed over to a test coordinator.

Failure to address incorrect, missing, or incomplete booklet information and labels may result in late reporting or no student score. In addition, the WIDA Consortium's national research agenda relies on complete and accurate student demographic data to inform the field and benefit English learners.

When preparing test materials for return to DRC, test administrators need to confirm that any booklet that contains student response information has either a Pre-ID Label or a District/School Label with bubbled student information. If a booklet is unused, there is no need to place any labels on the booklet. Placing a label on a booklet will cause it to be processed (and either scored, if the label is a Pre-ID or School/District label, or not scored, if it is a Do Not Process label).

6.3 *Rater Quality Control*

Rater Training: Students who take the ACCESS for ELLs Paper Speaking test have their spoken responses scored by the test administrator who administered the Speaking test. Another term for this test administrator is *rater*. Raters must be trained and certified, so we can be confident that they interpret students' spoken language consistently and fairly and that the scores are reported according to the WIDA English language proficiency standards. WIDA provides several different types of resources to support raters' training and reliability.

Students who take ACCESS for ELLs Online have their spoken responses digitally recorded and then scored centrally by DRC's trained raters. It is important that the individual who scores the spoken responses is trained and certified.

WIDA provides a series of training modules in the Secure Portal on the WIDA website. ACCESS for ELLs Speaking test raters should complete three core modules:

1. Overview and Test Structure
2. Speaking Assessment Scoring Practice
3. Speaking Assessment Recommended Practice

WIDA strongly recommends that all new raters complete all three of these modules. These modules provide a comprehensive introduction to the ACCESS for ELLs Speaking test and the opportunity to learn how to score students' spoken English reliably using the ACCESS for ELLs Speaking Scoring Scale.

In addition to the modules described above, WIDA also releases supplemental training materials each year to refamiliarize experienced raters with the Speaking Scoring Scale and introduce new Speaking tasks and sample responses for the coming year. These materials, called Supplemental Training for the Speaking Assessment, reflect the Speaking tasks that will appear on the test in the current year. WIDA recommends that all raters (new and experienced) engage with these supplementary materials at the start of each scoring season. Reading and reviewing these materials will help raters maintain their reliability from year to year and contribute to the fairness of test scores awarded to all students.

Rater Certification: After completing the training modules described above, new raters should take the relevant certification quiz. WIDA provides two quizzes: one for raters who will evaluate students in grades 1–5 and another for raters who will evaluate students in grades 6–12. Raters should take the appropriate quiz.

The purpose of the quiz is to ensure that raters have internalized the Speaking Scoring Scale and can apply it consistently. Only raters who pass the quiz(es) should administer and score the ACCESS for ELLs Paper Speaking test.

Checklist for Rater Training, Monitoring, and Recertification:

- New raters complete all Speaking Assessment Training
- New raters take and pass the appropriate certification quizzes
- All raters recertify at the start of each testing season (review new materials, retake quiz)
- Only certified raters administer and score the ACCESS for ELLs Speaking test
- Raters do not evaluate their own students, if at all possible
- Rater reliability and/or score point distributions are monitored regularly

For more information on Writing rater QC, please refer to Part 1, Section 4.2.

6.4 Score Reporting Quality Control

WIDA conducts an annual score reporting quality control process to (1) verify the accuracy of paper-based test scores (i.e., ACCESS for ELLs Paper, Kindergarten ACCESS for ELLs, and

Alternate ACCESS) and (2) verify the accuracy of all score reports (the Individual Student Report, the Student Roster Report, the School Frequency Report, the District Frequency Report, and the State Frequency Report) for both ACCESS (Online, Paper, and Kindergarten) and Alternate ACCESS.

The Score Reporting quality control is conducted at DRC's offices in Maple Grove, Minnesota. The team generally includes five state education agency representatives, one CAL employee, and four WIDA employees. This team examines data from three districts: a primary district, for quality control of all score reports; a secondary district, for quality control of State Frequency Reports only; and a tertiary district for quality control of paper-based tests only.

After an introductory presentation, which includes details of the quality control processes undertaken by DRC and WIDA and instructions on using the data entry tools, panelists begin by confirming the scoring of ACCESS Paper. Using the information in the State Student Response file, panelists enter the grade level, grade level cluster, tier, the Listening and Reading responses, and the Speaking and Writing scores into the data entry tool. The tool then calculates the student's raw scores and, using a series of look-ups, the student's scale score, proficiency level score, and confidence bands for all domains and composites. Panelists check student scores on the Individual Student Reports against those calculations. Any discrepancies are brought to the attention of the WIDA facilitator who investigates and, if there seems to be an issue with the report (rather than the data entry or data entry tool), discusses the issue further with DRC.

The panelists follow a similar process with the Kindergarten ACCESS tests, but with the raw scores for these tests copied directly from the response booklets.

After checking the paper-based tests, panelists turn their attention to the score reports. Panelists first check both the demographic information and the student scores in the Individual Student Reports against the information in the Student Roster Reports. Again, any discrepancies are brought to the attention of the facilitator, who investigates and discusses the issue with DRC if necessary. Panelists use the verified Individual Student Reports to check the Student Roster Report. Once the Student Roster Report is verified, panelists use it to check the State Frequency Report; they then use the verified State Frequency Reports to check the District Frequency Report. Finally, panelists check the State Frequency Reports against verified District Frequency Reports from the primary district along with District Frequency Reports from the secondary district.

6.5 Data Forensic Quality Control

Incidence of student plagiarism: DRC and WIDA have identified and confirmed instances of students plagiarizing responses of the Speaking and/or Writing tests for mostly clusters 68 and 912 items. While scoring student responses, DRC identified these students' responses as not being authentic to the student. WIDA staff have confirmed that students accessed the internet to look up specific wording from the task and to use information from a website in order to respond to the task. Some students produced spoken responses by utilizing an artificial voice (not the student's own voice), via either translation software or screen reading functionality.

When plagiarism was identified, the SEA representative in the state where the infraction occurred was notified immediately, and WIDA requested direction about those students' scores. All responses containing plagiarized content will receive a nonscorable code of "Invalid Indecipherable." This impacted 345 students in Speaking and 203 students in Writing across 36 states/territories.

Table 6.5.1 shows the summary of the number of students who plagiarized responses in the Speaking and/or Writing domains by state.

Table 6.5.1

Number of Plagiarisms

State	Speaking	Writing
AK*	1	0
AL*	2	0
CO*	0	3
DE*	1	1
FL*	0	2
GA*	11	12
HI*	4	0
ID*	4	4
IL*	55	41
IN*	20	7
KY*	4	1
MA*	10	6
MD*	23	11
ME*	5	0
MI	16	2
MN*	6	2
MO	7	7
MT*	2	5
NC*	28	11
ND*	5	2
NJ	11	1
NM*	10	9
NV*	9	9
OK	12	4
PA	21	17
RI*	5	4
SC*	2	0

State	Speaking	Writing
SD*	1	0
TN*	8	4
UT*	12	0
VA*	14	16
VI*	3	6
VT*	0	1
WA*	23	10
WI*	9	5
WY*	1	0
Total	345	203

* = states where scoring is complete and all flagged suspected K plagiarisms have been reported to SEA
 Note: Counts represent # of students that were flagged for suspected plagiarisms. Some students were flagged for multiple responses, so overall response count flagged is higher.

Suspected AI-generated Responses: On January 30th, 2024, the DRC scoring team noticed several speaking responses that were suspected of being generated via AI tools. A suspected AI-generated response is evidenced by students reading from scripts or an external resource; however, the external source cannot be identified with a direct website link for reference. The response may sound unnatural and contain detailed information or technical vocabulary that is not provided in the task input and not likely for students to know offhand.

WIDA and DRC worked on an iterative process for flagging and reviewing the suspected AI-generated responses for further investigation. All suspected AI-generated responses are reviewed by scoring supervisors. They are scored as usual but internally flagged with a tag. These responses are counted toward student scores. The DRC scoring team uploads suspected AI-generated responses as flagged, and the WIDA content team conducts an independent review of responses with transcripts and comments. The WIDA team confirms that these responses are not original language produced by the student but are read aloud from some source text. However, these sources cannot be directly referenced to a website or an external source, and the responses could not be fully replicated. States are advised to conduct further review and investigation.

DRC provides flagged responses without identifying student information to WIDA for its independent review. However, DRC includes student information when providing suspected AI-generated responses to the states for further review and investigation. The DRC scoring team flagged a total of 448 suspected AI-generated responses from 352 students (excluding duplicates; some students were flagged for multiple responses). The final count includes 308 speaking responses from 222 students and 140 writing responses from 130 students (34 states).

Suspected Item Exposure: Between October 5, 2023, and June 16, 2024, WIDA, state partners, and Caveon identified 53 posts on social media or other websites containing ACCESS-related content, out of which 14 were related to sample items or practice materials,

and 1 included a retired item. Thirty-six of the posts included operational items, and 1 included a field test item. On one post, we were unable to determine the item's status due to the constraints of the social media platform.

Across the 36 posts that included operational content, the following number of items were exposed:

Listening 9–12: 3 items

Reading 9–12: 1 item

Speaking 6–8: 3 items

Speaking 9–12: 3 items

Writing 45: 1 item

Writing 9–12: 2 items

All posts were removed from social media upon request.

An item is suspected of being exposed if any content appears on social media. The WIDA test development team reviewed images and videos to identify the exact screens that clearly contained content related to tasks, prompts, and response options. The WIDA psychometrics team conducted analyses comparing item performance before and after items were exposed against overall item performance. Item parameters from the previous testing year were compared against this year's item parameters using the data with potential item exposure. WIDA also reviewed and compared item statistics before and after the items appeared on social media. Given that these posts were promptly removed from local devices or social media, the results suggested little variation regarding item performance. WIDA has decided to retain operational items for scoring, but exposed items were excluded from item calibration for verification studies for operational items and will not appear on future test administrations. Any field test items that were exposed will not be part of next year's operational test.

Caveon Data Forensic Analysis Results: WIDA hired Caveon to perform data forensic analysis during the 2023–2024 test administration cycle to examine whether ACCESS data has been compromised or has evidence of item exposure.

Caveon security statistics are based on mathematical models, where the test response data are used to create a baseline model of normal or "typical" test taking among that population. Individuals or groups are then compared to the baseline, and observations that are significantly different from the baseline are flagged as anomalous. Caveon's statistics are designed to be robust but also conservative regarding which and how many individuals or groups are flagged as anomalous, thereby reducing the chances of false-positive detections.

Data forensics analysis was performed after the administration window for the following administrations:

- December 2023 through August 2024 online multistage adaptive test administrations, Listening and Reading domains
- December 2023 through August 2024 paper fixed-form administrations, Listening and Reading domains

The analysis utilized several of Caveon's security statistics to detect evidence of whether the assessment instrument has been compromised through disclosure of the content. This analysis attempted to understand where and when disclosure of the test content may have occurred and what items and forms may have been affected. Results of this analysis might enable WIDA to take specific actions to limit the impact of disclosed content. Such actions may include

- Republishing or reworking items or forms
- Rotating disclosed items to limit their exposure
- Designing a republication or rotation strategy for future items and forms

Caveon security statistics were computed for each individual test instance. These data were aggregated or summarized at the group level. The aggregated statistics were compared against the population model.

Analysis of Tests: Caveon aggregated the data according to individual test forms using the security statistics to determine whether rates of detections by the security statistics were higher for certain test forms. For fixed-form paper tests, two forms—A and B/C—were analyzed. For the multistage adaptive test, there is a finite number of ways a student could progress through the test. Caveon analyzed each pathway as a separate form. Higher rates of security detections for a specific form of the test suggest that compromise of the form may have occurred.

Analysis of Items:

Item security: In this portion of the analysis, the security of the items was evaluated using aberrance statistics. Aberrance statistics detect test-taking behaviors such as answering difficult items correctly but answering easy items incorrectly, or unusual patterns in the time taken to answer test items. In the absence of security issues, aberrant test taking is expected to be the result of poor or uneven test preparation, illness or other physical malady, mental and emotional distractions, and so forth. These factors usually result in lower levels of test performance. When aberrance is associated with higher performance, however, test fraud may have occurred, such as preknowledge of test content. By applying aberrance measures and comparing the performance between aberrant and nonaberrant test instances on individual items, inferences can be made about item security.

Item performance changes: Analysis of item performance changes tracks individual item performance rates over time. The item performance shifts are measured within the context of the item response theory model and adjusted for varying test-taker performance levels. This means that detected performance shifts are invariant to fluctuations in the test-taker population. When performance shifts indicate the item has become significantly easier, the item may have been disclosed. Items with significant performance shifts become candidates for revision or replacement. Item performance shifts were detected with a granularity of 1 week, where Monday to Sunday represents 1 week.

Analysis of Groups:

Analysis by week: This analysis aggregates the data according to the week in which the test was taken to identify whether security threats and pass rates appeared to be more prevalent at certain times during the testing window. Increases in scores or security detections during

certain periods of time suggest the content may have been disclosed at some point prior to that time. This analysis also includes a form-date grouping to determine if increasing security threats are associated with a particular form of the test. This analysis is performed for online and paper tests, where relevant test date data are provided.

Analysis of WIDA jurisdictions: Caveon analyzed WIDA member jurisdictions (states and districts) to determine whether rates of detections by the security statistics were higher for certain jurisdictions. This analysis is intended to detect whether compromise at the state or member jurisdiction level potentially occurred. This analysis is performed for online and paper tests.

Analysis of administration mode: Caveon aggregates the data according to administration mode (i.e., online versus paper) to determine if security threats are associated with the mode of testing.

Other Analyses:

Analysis of mean score over time: Analysis of mean score over time was used to identify whether mean scores increased over time during the testing window. Increases in scores over time suggest the content may have been disclosed during the testing window.

Findings of Data Forensic Analyses: Generally, no major data forensic anomalies were observed across WIDA states. A few minor localized anomalies associated with items are under WIDA's investigation.