Background

In 2014, student performance on the Grade-level Missouri Assessment Program (MAP) fell from 2013. In response to this, the Missouri Department of Elementary and Secondary Education (DESE) contracted with the National Center for the Improvement of Educational Assessment, Inc. (Center for Assessment) to examine the 2014 administration of the Grade-level MAP to look for possible causes for the drop in scale scores and percentage of students at or above Proficient from 2013.

Tests should be administered under similar conditions, at the same time of year, and to similar populations. In this paper, we describe the results of our examination of various factors related to the assessment program that may have influenced the results of the test. These factors include the comparability of the:

- Test population
- Test construct
- Test administration
- Psychometric characteristics
- Internal anchor

Prior to discussing each of these factors, we first examine the test results that precipitated this work and then explain the data that DESE provided for these analyses. All analyses in this paper were conducted between July 28 to August 1. During the window, the crux of the analyses focused on those grades where the decrease in test performance was largest (Grades 3 and 4 Communication Arts and Grade 4 Mathematics).

Test Results

In 2014, the results on the grade-level MAP fell from previous performance. These decreases in test performance were especially large in Grades 3 and 4 Communication Arts and Grade 4 Mathematics.

Table 1 aggregates the grade levels by the decrease or gain between 2013 and 2014 in the percentage at or above Proficient.

- Large decrease= Percent at or above Proficient in 2014 decreased by more than 5 percentage points from 2013.
- Moderate decrease = Percent at or above Proficient in 2014 decreased by 3 to 5 percentage points from 2013.
- Small decrease = Percent at or above Proficient in 2014 decreased by 1 to 3 percentage points from 2013.
- No change = Percent at or above Proficient in 2014 was within +/- .5 percentage point of 2013.

- Gain=Percent at or above Proficient increased by 1 or more percentage points from the previous years. (The largest gain occurred in Grade 8 math where at or above Proficient increase by 2.5 percentage points.)

Table 1 Grade Levels Aggregated by Drop/Gain in Percentage At or Above Proficient

Content Area	Large Decrease	Moderate Decrease	Small Decrease	No Change	Gain
Communication Arts	3, 4	6	5		7
Mathematics	4	8	5	3, 6, 7	8
Science		5			8

As shown in Table 1, the largest drops were observed in:

- Grade 3 Communication Arts
- Grade 4 Communication Art
- Grade 4 Mathematics

Tables 2 through 5 show detailed results for all grades and content area. Table 2 shows the n-counts, mean scale scores, and standard deviation of scale scores for all years of the testing program. Tables 3, 4, and 5 show the percentage of students in each of achievement levels at each grade for Communication Arts, Mathematics, and Science, respectively.

The results for 2006 through 2013 were taken from the Technical Report of the 2013 Grade-Level Missouri Assessment Program

Table 2 Mean Scale Scores for Communication Arts, Mathematics, and Science, 2006-2014

		Communication Arts			Mathematics			Science		
	Year	N	Mean	SD	N	Mean	SD	N	Mean	SD
	2006	64,486	639.86	36.84	64,763	621.59	39.11			
	2007	66,347	639.58	38.04	66,640	622.4	38.72			
	2008	66,179	637.6	37.54	66,258	621.65	36.92			
3	2009	67,163	637.43	38.18	67,232	621.67	36.76			
	2010	66,751	640.27	36.63	66,814	624.89	39.28			
	2011	66,196	641.19	36.52	66,258	627.03	39.69			
	2012	66,147	641.78	37.66	66,213	628.65	39.78			
	2013	66,491	643.69	37.18	66,609	627.92	39.54			
	2014	67,028	638.29	39.49	67,080	628.00	41.73			
	2006	65,179	654.55	38.56	65,306	643.88	37.07			
	2007	65,274	656.11	39.51	65,363	644.47	36.56			
	2008	66,873	655.61	33.63	66,944	644.18	34.19			
4	2009	66,490	656.77	33.41	66,587	644.2	33.89			
	2010	67,301	661.34	38.95	67,394	647.59	34.01			
	2011	66,748	662.18	38.23	66,881	649.68	34.87			
	2012	65,828	662.31	39.33	65,909	649.36	34.88			
	2013	65,859	662.83	38.85	65,991	648.98	33.84			
	2014	66,454	657.49	35.64	66,535	644.27	34.32			
	2006	66,007	668.18	37.09	66,123	660.06	39.99			
	2007	65,461	671.01	37.14	65,498	663.21	41.5			
	2008	65,544	671.48	33.71	65,636	661.43	40.73	65,586	661.64	31.52
5	2009	67,083	671.58	32.84	67,155	662.07	40.52	67,118	662.22	30.4
	2010	66,500	673.65	35.33	66,580	667.7	41.74	66,558	664.76	32.48
	2011	67,052	673.68	34.85	67,124	669.05	42.48	67,196	666.04	33.43
	2012	66,470	674.16	35.44	66,524	670.61	42.80	66,492	667.99	34.23
	2013	65,714	674.7	35.14	65,861	670.18	42.84	65,846	667.55	32.99
	2014	65,959	673.19	34.66	66,021	667.94	43.81	66,000	664.06	30.51
	2006	66,948	666.85	33.7	67,017	673.3	39.8			
	2007	66,247	667.99	34.63	66,332	676.31	41.75			
	2008	65,672	671.27	33.5	65,716	678.46	41.13			
6	2009	65,716	671.67	33.04	65,755	678.87	39.56			
	2010	67,260	674.18	33.12	67,315	683.36	39.48			
	2011	66,443	675.02	32.81	66,476	684.95	39.8			
	2012	67,173	674.33	32.83	67,237	684.43	40.19			
	2013	66,430	675.06	32.33	66,509	685.02	39.85			
	2014	65,785	672.46	33.46	65,806	683.86	39.73			

		Communication Arts			Mathematics			Science		
	Year	N	Mean	SD	N	Mean	SD	N	Mean	SD
	2006	70,290	671.63	37.06	70,698	675.38	41.27			
	2007	67,167	672.11	36.26	67,554	677.41	42.62			
	2008	66,701	675.87	35.08	66,727	681.15	41.38			
7	2009	66,316	677.68	34.75	66,330	683.63	40.72			
	2010	66,034	678.85	36.25	66,052	686.51	40.28			
	2011	67,257	680.56	36.61	67,294	687.53	40.73			
	2012	66,620	681.73	36.19	66,654	691.18	41.51			
	2013	67,065	681.39	36.11	66,300	689.65	41.24			
	2014	66,614	681.95	35.97	65,789	689.78	40.55			
	2006	72,483	686.85	37.87	72,542	697.73	40.37			
	2007	70,187	686.9	37.54	70,204	698.33	41.98			
	2008	67,278	691.05	33.57	67,312	701.3	39.4	67,209	694.36	30.67
8	2009	66,741	692.56	33.31	66,770	703.6	38.63	66,702	695.65	30.94
	2010	66,139	694.28	34.01	66,166	707.98	40.04	66,101	698.28	31.07
	2011	65,905	695.11	34.1	65,956	708.4	40.12	65,828	700.05	30.98
	2012	66,755	695.89	33.52	66,808	709.57	40.2	66,724	700.18	31.92
	2013	66,349	696.34	32.95	51,570*	699.9	35.69	66,414	699.93	31.68
	2014	66,834	693.84	31.54	52,997	700.22	37.98	66,893	701.94	29.54

Table 3 Percent in Each Achievement Level, Communication Arts 2006-2014

Grade	Year	N	No Level	Below Basic	Basic	Proficient	Advanced	Prof & Adv	Difference (Year-Prev Year)
	2006	65,344	1.3	8.8	47.5	25.7	16.7	42.4	
	2007	67,259	1.4	9.4	46.6	25.8	16.8	42.6	0.2
	2008	66,357	0.3	9.3	50.2	25.2	15.1	40.3	-2.3
3	2009	67,357	0.3	9.6	49.8	25.1	15.2	40.3	0
	2010	66,947	0.3	8.2	48.4	26.9	16.2	43.1	2.8
	2011	66,487	0.4	7.6	48.4	27	16.6	43.6	0.5
	2012	66,323	0.3	8	46.5	27.2	18.1	45.3	1.7
	2013	66,754	0.3	7.8	44.2	27.7	20.1	47.8	2.5
	2014	67,028		9.8	48.5	25.6	16.1	41.7	-6.1
	2006	65,849	1	10.6	44.5	28.8	15	43.8	
	2007	65,982	1.1	10.5	43.4	28.2	16.8	45.1	1.3
	2008	67,049	0.3	8	46.7	33.4	11.7	45.1	0
4	2009	66,709	0.3	7.6	45.8	33.6	12.7	46.3	1.2
	2010	67,510	0.3	8.6	40.2	31.2	19.7	50.9	4.6
	2011	67,049	0.4	8.2	39.5	31.6	20.2	51.9	1
	2012	65,996	0.3	8.3	39.3	31.2	20.9	52.2	0.3
	2013	66,085	0.3	8.2	38.7	31.6	21.2	52.8	0.6
	2014	66,454		7.8	46.5	31.6	14.1	45.7	-7.1
	2006	66,704	1	9.1	44.8	29.6	15.4	45	
	2007	66,098	1	8.3	42.9	29.8	18	47.8	2.8
	2008	65,734	0.3	6.4	45.1	32.2	15.9	48.1	0.3
5	2009	67,307	0.3	6.3	44.6	33.9	14.9	48.8	0.7
	2010	66,730	0.3	7.1	41.5	32.1	18.9	51	2.2
	2011	67,461	0.6	6.9	41.4	32.4	18.7	51.1	0.1
	2012	66,675	0.3	7	40.9	32.3	19.6	51.8	0.7
	2013	65,980	0.3	7.1	40.3	32.2	20.1	52.3	0.5
	2014	65,959		6.3	43.6	33.3	16.9	50.2	-2.1

Grade	Year	N	No Level	Below Basic	Basic	Proficient	Advanced	Prof & Adv	Difference (Year-Prev Year)
	2006	67,709	1.1	11.9	44.8	31.6	10.6	42.2	
	2007	67,045	1.2	11.2	44	31.8	11.7	43.6	1.4
	2008	65,830	0.2	9	43.5	34	13.4	47.4	3.8
6	2009	65,908	0.3	8.6	43.4	33.8	13.9	47.7	0.3
	2010	67,476	0.3	7.8	42.3	33.9	15.7	49.6	1.9
	2011	66,633	0.3	7.3	41.9	34.3	16.2	50.5	0.9
	2012	67,342	0.3	7.5	42	34.7	15.5	50.2	-0.3
	2013	66,731	0.4	7.2	41.4	34.9	16.1	51	0.8
	2014	65,785		8.5	43.9	33	14.6	47.6	-3.4
	2006	71,632	1.9	13.7	41.8	30.5	12.2	42.7	
	2007	68,404	1.8	13.1	40.7	32.8	11.6	44.4	1.7
	2008	66,923	0.3	10	40.7	36.1	12.9	49	4.6
7	2009	66,531	0.3	8.7	40.3	37.2	13.6	50.8	1.8
	2010	66,279	0.4	9.8	38.1	35.2	16.5	51.7	0.9
	2011	67,517	0.4	9	36.9	36	17.8	53.8	2.1
	2012	66,845	0.3	8.7	35.8	36.6	18.7	55.2	1.4
	2013	67,319	0.3	9	35.7	36.5	18.4	54.9	-0.3
	2014	66,614		8.2	36.1	37	18.6	55.6	0.7
	2006	73,516	1.4	9.1	48	26.6	15	41.5	
	2007	71,200	1.4	8.7	48.3	26.9	14.6	41.6	0.1
	2008	67,574	0.4	5.7	45.8	33.1	15	48.1	6.5
8	2009	67,077	0.5	5.3	44.5	33.4	16.3	49.7	1.6
	2010	66,463	0.5	4.9	42.8	34.3	17.4	51.8	2.1
	2011	66,205	0.5	4.6	42.5	33.9	18.5	52.5	0.7
	2012	67,037	0.4	4.3	42	34.3	19	53.3	0.8
	2013	66,710	0.5	4.1	41.5	34.9	19	53.9	0.6
	2014	66,834		4.5	44.9	34.3	16.4	50.7	-3.2

Table 4 Percent in Each Achievement Level, Mathematics 2006-2014

Grade	Year	N	No Level	Below Basic	Basic	Proficient	Advanced	Prof & Adv	Difference (Year-Prev Year)
	2006	65,325	0.9	7.2	48.7	33.3	10	43.3	
	2007	67,257	0.9	7.2	46.9	35	10	45	1.7
	2008	66,357	0.1	6.5	49.6	35	8.8	43.8	-1.2
3	2009	67,357	0.2	6.8	48.5	35.6	8.8	44.4	0.6
	2010	66,947	0.2	6.2	46.6	37	10.1	47.1	2.7
	2011	66,487	0.3	5.6	44.7	38.1	11.3	49.4	2.3
	2012	66,323	0.2	5.4	42.6	39.9	11.9	51.9	2.5
	2013	66,754	0.2	5.3	43.8	39.2	11.4	50.7	-1.2
	2014	67,080		6.0	43.7	36.7	13.6	50.3	-0.4
	2006	65,845	0.8	8.3	47.5	34.4	9	43.4	
	2007	65,975	0.9	8.1	46.5	35.2	9.3	44.5	1.1
	2008	67,049	0.2	7.6	48	36	8.2	44.2	-0.3
4	2009	66,709	0.2	7.3	48.2	36.6	7.8	44.4	0.2
	2010	67,510	0.2	6.1	45.4	39.3	9.1	48.4	4
	2011	67,049	0.3	5.6	43.7	39.9	10.5	50.5	2.1
	2012	65,996	0.1	5.7	43.7	40.5	10	50.5	0
	2013	66,085	0.1	5.5	44.2	40.7	9.4	50.1	-0.4
	2014	66,535		6.6	51.2	34.6	7.6	42.2	-7.9
	2006	66,703	0.9	8.1	47.8	32.7	10.6	43.3	
	2007	66,075	0.9	7.6	44.9	33.1	13.4	46.6	3.3
	2008	65,734	0.1	7.5	46.5	34.4	11.4	45.8	-0.8
5	2009	67,307	0.2	7.5	45.1	35.6	11.6	47.2	1.4
	2010	66,730	0.2	6.2	41.9	36.7	15.1	51.7	4.5
	2011	67,461	0.5	6.1	40.9	36.3	16.2	52.5	0.8
	2012	66,675	0.2	5.8	39.7	35.9	18.4	54.3	1.8
	2013	65,980	0.2	5.9	40.1	35.9	18	53.9	-0.4
	2014	66,021		7.2	40.6	35.5	16.7	52.2	-1.7

Grade	Year	N	No Level	Below Basic	Basic	Proficient	Advanced	Prof & Adv	Difference (Year-Prev Year)
	2006	67,706	1	11.1	44.1	34.4	9.5	43.9	
	2007	67,039	1.1	11.1	40	35.5	12.3	47.8	3.9
	2008	65,830	0.2	9.5	39.6	37.8	12.9	50.7	2.9
6	2009	65,908	0.2	8.9	40.7	37.5	12.6	50.1	-0.6
	2010	67,476	0.2	7.8	36.6	40.3	15	55.4	5.3
	2011	66,633	0.2	7.5	35.4	40.5	16.4	56.9	1.5
	2012	67,342	0.2	7.4	36.7	39.7	16	55.7	-1.2
	2013	66,731	0.3	7	36.4	39.9	16.3	56.2	0.5
	2014	65,806		7.2	37	40.5	15.3	55.8	-0.4
	2006	71,575	1.2	17.4	38.5	32.7	10.2	42.9	
	2007	68,405	1.2	16.7	37.1	33.2	11.7	44.9	2
	2008	66,923	0.3	13.9	36.3	36.7	12.8	49.5	4.6
7	2009	66,531	0.3	12.5	35.2	37.6	14.3	51.9	2.4
	2010	66,279	0.3	10.8	34.3	38.8	15.7	54.5	2.6
	2011	67,517	0.3	10.5	33.5	39.2	16.6	55.8	1.3
	2012	66,845	0.3	9.8	30.3	40	19.6	59.6	3.8
	2013	67,319	1.5	10.1	31.1	39.1	18.2	57.3	-2.3
	2014	65,789		9.8	32.5	39.2	18.5	57.7	0.4
	2006	73,523	1.3	21.1	37.8	27.6	12.2	39.8	
	2007	71,190	1.4	21.4	36.6	26.6	14	40.6	0.8
	2008	67,574	0.4	18	37.7	29.9	13.9	43.8	3.2
8	2009	67,077	0.5	16.4	36.8	31.5	14.9	46.4	2.6
	2010	66,463	0.4	14.9	33.3	32.1	19.2	51.3	4.9
	2011	66,205	0.4	15	33.9	31	19.8	50.8	-0.5
	2012	67,037	0.3	14.1	33.6	31.8	20.2	52	1.2
	2013	52,335*	1.4	17.1	41.2	30.2	10.1	40.3	-11.7
	2014	51,997		17.9	39.3	31.3	11.5	42.8	2.5

Table 5 Percent in Each Achievement Level, Science 2006-2014

Grade	Year	N	No Level	Below Basic	Basic	Proficient	Advanced	Prof & Adv	Difference (Year-Prev Year)
	2008	65,734	0.2	11.2	44	29.6	14.9	44.5	
	2009	67,307	0.3	10.6	44.1	30.3	14.8	45.1	0.6
	2010	66,730	0.3	10.4	40.5	29.6	19.3	48.9	3.8
5	2011	67,461	0.4	10	39.1	29.5	21	50.5	1.6
	2012	66,675	0.3	9.8	38.5	27.2	24.3	51.4	0.9
	2013	65,980	0.2	9.6	39	28.1	23.1	51.3	-0.1
	2014	66,000		9.0	43.5	31.6	15.9	47.5	-3.8
	2008	67,574	0.5	19.3	37	36.7	6.5	43.2	
	2009	67,077	0.6	18.2	36.5	37.2	7.6	44.8	1.6
	2010	66,463	0.5	16.4	35.1	38.4	9.6	48	3.2
8	2011	66,205	0.6	15.7	33.7	38.6	11.4	50	2.0
	2012	67,037	0.5	16.1	33.8	37	12.6	49.6	-0.4
	2013	66,710	0.4	15.7	33.8	38.4	11.6	50.1	0.5
	2014	66,893		12.8	35.1	40.6	11.4	52.0	1.9

Data

DESE provided several data files and psychometric output in order to conduct this analysis.

Statewide Testing Data

DESE provided student-level statewide data files for the 2013 and 2014 test. The following variables were included in these data files:

- District identifier
- School identifier
- Demographic variables
  - Gender, race/ethnicity, free and reduced lunch
- Testing variables
  - Content area, grade, achievement level, MAP score, item-level scores, content-standard scores, process-standard scores, invalidation flag

No personally identifiable information was included in the student-level data. This meant that no student-level, longitudinal analyses could be conducted.

Statewide District Information

DESE provided separate district-level Excel files for each of the following:

- Spring Break information
- Pick-up dates (shows the date the tests were picked up by the vendor)

- Participation in the Smarter Balanced field test
- Number of school days completed, number of snow days

Psychometric Output

DESE provided CTB/McGraw-Hill's output files of item analysis, calibration, and Stocking & Lord equating of the 2014 assessment. DESE provided files with p-values for anchor items from their original administration and form the 2014 administration.

In addition, DESE provided files of CTB's anchor selection conducted in Spring 2013.

Item Metadata

DESE provided item metadata files from testing years 2006 through 2014. These files provide detailed information on each item, including item's placement on the test (item and page), content measured, item type, and anchor/non-anchor status.

Comparability of Testing Population

In order to examine the comparability of the test population, frequency distributions of demographic variables from 2014 was compared to the distribution of the same variables in 2013. The purpose of these comparisons was to examine the stability of the testing population in 2014 compared to 2013. If the number and types of students has changed dramatically since 2013, then this may have contributed to the decrease in scores.

The total n-counts are reported in Table 2, Table 3, Table 4, and Table 5. As can be seen in these tables, the total n-counts are stable from year to year. The n-counts from 2014 are well within the same range as 2006 to 2013.

Table 6 through Table 9 shows detailed information comparing frequency distributions from 2014 and 2013 for gender (Table 6), race/ethnicity (Table 7), IEP assignment (Table 8), and free/reduced lunch (Table 9). The distribution of students within each variable in 2014 is very similar to the distribution of students within the same variable in 2014. For almost all comparisons, there is less than a one percentage point difference between 2014 and 2013. In one only case, the percentage of student receiving free or reduced lunch, was there a larger than one percentage point difference. In Grade 3, the percentage of students receiving free or reduced lunch increased 1.2 percentage points from 2013.

Tables 6 through 9 also compare total n-counts. These show that the n-counts changed by no more than 700 students (approximately 1%) from the previous year.

Summary of Testing Population Comparability

The information in Tables 6 through 9 indicates that the 2014 population is comparable to the 2013 population. These analyses indicate that the decrease in scores from 2013 to 2014 was not due to a change in the student population.

Table 6 Percent of Students in each Grade by Content Area Disaggregated by Gender, 2014 and 2013, 2014 and 2013

		Communication Arts			Mathematics			Science		
Grade	Gender	2014	2013	Difference	2014	2013	Difference	2014	2013	Difference
3	Missing	.0	.0	0.02	.0	.0	0.02			
	Female	48.9	48.8	0.15	48.9	48.8	0.15			
	Male	51.0	51.2	-0.17	51.0	51.2	-0.17			
	Total N-Count	67028	66547	481.00	67080	66611	469.00			
4	Missing	.0	.0	0.02	.0	.0	0.02			
	Female	48.7	48.8	-0.05	48.7	48.8	-0.05			
	Male	51.2	51.2	0.03	51.2	51.2	0.03			
	Total N-Count	66454	65883	571.00	66535	65975	560.00			
5	Missing	.0	.0	0.01	.0	.0	0.01	.0	.0	0.01
	Female	48.8	48.7	0.05	48.8	48.7	0.04	48.8	48.7	0.04
	Male	51.2	51.3	-0.07	51.2	51.3	-0.05	51.2	51.3	-0.06
	Total N-Count	65959	65770	189.00	66021	65855	166.00	66000	65844	156.00
6	Missing	.0	.0	0.00	.0	.0	0.00			
	Female	48.7	49.3	-0.62	48.7	49.3	-0.64			
	Male	51.3	50.7	0.62	51.3	50.7	0.64			
	Total N-Count	65785	66480	-695.00	65806	66497	-691.00			
7	Missing	.0	.0	0.00	.0	.0	0.00			
	Female	49.3	48.5	0.79	49.4	48.5	0.85			
	Male	50.7	51.5	-0.79	50.6	51.4	-0.85			
	Total N-Count	66614	67158	-544.00	65789	66362	-573.00			
8	Missing	.0	.0	0.00	.0	.0	0.01	.0	.0	0.00
	Female	48.6	48.8	-0.17	47.5	47.8	-0.31	48.6	48.8	-0.18
	Male	51.3	51.2	0.16	52.4	52.1	0.30	51.4	51.2	0.17
	Total N-Count	65789	66362	-573.00	100.0	100.0	0.00	66893	66464	429.00

Table 7 Percent of Students in each Grade by Content Area Disaggregated by Race/Ethnicity Code, 2014 and 2013

Grade	Race/Ethnicity Code	Communication Arts			Mathematics			Science		
		2014	2013	Difference	2014	2013	Difference	2014	2013	Difference
3	0	.4	.4	0.00	.4	.4	0.00			
	1	2.0	1.9	0.02	2.0	2.0	0.03			
	2	.2	.2	0.04	.2	.2	0.04			
	3	16.1	16.6	-0.45	16.1	16.6	-0.47			
	4	6.1	5.7	0.40	6.1	5.7	0.40			
	5	72.0	72.9	-0.86	71.9	72.8	-0.85			
	6	3.1	2.3	0.82	3.1	2.3	0.82			
	Total Valid	67002	66535	467.00	67055	66599	456.00			
	Missing	.0	.0	0.02	.0	.0	0.02			
	Total N-Count	67028	66547	481.00	67080	66611	469.00			
4	0	.5	.4	0.04	.5	.4	0.04			
	1	1.9	2.1	-0.12	2.0	2.1	-0.12			
	2	.2	.2	0.02	.2	.2	0.02			
	3	16.3	16.2	0.07	16.3	16.2	0.06			
	4	5.7	5.5	0.21	5.8	5.6	0.21			
	5	72.8	73.3	-0.55	72.7	73.3	-0.56			
	6	2.5	2.2	0.32	2.5	2.2	0.32			
	Total Valid	66437	65880	557.00	66518	65971	547.00			
	Missing	.0	.0	0.02	.0	.0	0.02			
	Total N-Count	66454	65883	571.00	66535	65975	560.00			

Grade	Race/Ethnicity Code	Communication Arts			Mathematics			Science		
		2014	2013	Difference	2014	2013	Difference	2014	2013	Difference
5	0	.4	.5	-0.05	.4	.5	-0.06	.4	.5	-0.06
	1	2.1	2.0	0.07	2.1	2.0	0.07	2.1	2.0	0.07
	2	.2	.2	0.01	.2	.2	0.01	.2	.2	0.01
	3	16.1	16.5	-0.37	16.1	16.5	-0.39	16.1	16.5	-0.39
	4	5.6	5.4	0.15	5.6	5.5	0.16	5.6	5.5	0.16
	5	73.2	73.3	-0.11	73.1	73.2	-0.10	73.1	73.2	-0.10
	6	2.4	2.1	0.29	2.4	2.1	0.29	2.4	2.1	0.29
	Total Valid	65942	65762	180.00	66004	65845	159.00	65983	65836	147.00
	Missing	.0	.0	0.01	.0	.0	0.01	.0	.0	0.01
	Total N-Count	65959	65770	189.00	66021	65855	166.00	66000	65844	156.00
6	0	.4	.4	0.02	.4	.4	0.02			
	1	2.0	1.8	0.14	2.0	1.8	0.16			
	2	.2	.2	-0.03	.2	.2	-0.03			
	3	16.3	16.3	0.00	16.3	16.3	0.01			
	4	5.5	5.1	0.38	5.6	5.2	0.37			
	5	73.3	74.1	-0.83	73.2	74.0	-0.85			
	6	2.3	2.0	0.31	2.3	2.0	0.31			
	Total Valid	65768	66463	-695.00	65789	66481	-692.00			
	Missing	.0	.0	0.00	.0	.0	0.00			
	Total N-Count	65785	66480	-695.00	65806	66497	-691.00			

Grade	Race/Ethnicity Code	Communication Arts			Mathematics			Science		
		2014	2013	Difference	2014	2013	Difference	2014	2013	Difference
7	0	.4	.4	-0.03	.4	.4	-0.03			
	1	1.8	1.9	-0.02	1.7	1.7	-0.02			
	2	.2	.1	0.04	.2	.1	0.05			
	3	16.2	16.6	-0.41	16.3	16.7	-0.34			
	4	5.3	4.9	0.43	5.4	4.9	0.44			
	5	73.9	74.2	-0.29	73.8	74.2	-0.38			
	6	2.2	1.9	0.27	2.1	1.9	0.28			
	Total Valid	66602	67147	-545.00	65777	66351	-574.00			
	Missing	.0	.0	0.00	.0	.0	0.00			
	Total N-Count	66614	67158	-544.00	65789	66362	-573.00			
8	0	.5	.5	-0.03	.5	.5	-0.03	.5	.5	-0.03
	1	1.9	1.7	0.18	1.3	1.2	0.16	1.9	1.7	0.18
	2	.1	.2	-0.02	.2	.2	-0.01	.2	.2	-0.02
	3	16.5	16.7	-0.17	18.2	18.5	-0.22	16.5	16.6	-0.14
	4	4.9	4.6	0.33	5.3	4.7	0.58	5.0	4.6	0.37
	5	74.2	74.7	-0.51	72.5	73.3	-0.72	74.1	74.6	-0.58
	6	1.9	1.7	0.22	2.0	1.7	0.23	1.9	1.7	0.21
	Total Valid	66825	66436	389.00	51988	51742	246.00	66884	66457	427.00
	Missing	.0	.0	0.00	.0	.0	0.01	.0	.0	0.00
	Total N-Count	66834	66443	391.00	51997	51748	249.00	66893	66464	429.00

Table 8 Percent of Students in each Grade by Content Area and IEP assignment, 2014 and 2013

		Communication Arts			Mathematics			Science		
GRADE	IEP	2014	2013	Difference	2014	2013	Difference	2014	2013	Difference
3	Yes	12.8	12.9	-0.08	12.8	12.8	-0.08			
	Total N-Count	67028	66547	481.00	67028	66547	481.00			
4	Yes	13.1	12.9	0.17	13.1	12.9	0.17			
	Total N-Count	66454	65883	571.00	66535	65975	560.00			
5	Yes	12.7	12.6	0.08	12.6	12.6	0.07	12.6	12.6	0.07
	Total N-Count	65959	65770	189.00	66021	65855	166.00	66000	65844	156.00
6	Yes	12.0	12.1	-0.12	12.0	12.1	-0.12			
	Total N-Count	65785	66480	-695.00	65806	66497	-691.00			
7	Yes	11.5	11.6	-0.12	11.6	11.7	-0.11			
	Total N-Count	66614	67158	-544.00	65789	66362	-573.00			
8	Yes	11.2	11.2	-0.04	13.8	13.8	-0.04	11.1	11.2	-0.06
	Total N-Count	66834	66443	391.00	51997	51748	249.00	66893	66464	429.00

Table 9 Percent of Students in each Grade by Content Area and Participation in the Free & Reduced Lunch Program, 2014 and 2013

		Communication Arts			Mathematics			Science		
Grade	Free & Reduced Lunch	2014	2013	Difference	2014	2013	Difference	2014	2013	Difference
3	Yes	54.6	53.4	1.2	54.6	53.4	1.2			
	Total N-Count	67028	66547	481	67080	66611	469			
4	Yes	52.9	52.8	0.1	52.9	52.8	0.1			
	Total N-Count	66454	65883	571	66535	65975	560			
5	Yes	52.2	52.4	-0.2	52.2	52.4	-0.2	52.2	52.4	-0.2
	Total N-Count	65968	65770	198	66030	65855	175	66009	65844	165
6	Yes	51.6	51.3	0.3	51.6	51.4	0.2			
	Total N-Count	65803	66480	-677	65824	66497	-673			
7	Yes	50.3	49.8	0.5	50.8	50.2	0.6			
	Total N-Count	66640	67158	-518	65815	66362	-547			
8	Yes	48.8	48.3	0.5	54.8	54.3	0.5	48.8	48.3	0.5
	Total N-Count	66849	66443	406	52012	51748	264	66908	66464	444

Comparability of Administration Conditions

In the best-case scenario, the administration of the test should be similar from year to year. In other words, annual summative tests should be administered under similar conditions, at the same time of year.

In 2014, there were two changes in administration conditions from 2013 that may have impacted student performance on the state assessment. First, there were a large number of snow days across the state during the 2013-2014 school year. Districts who missed more school due to snow may have decreases in performance from 2013 to 2014 due to lost instructional time and disrupted scheduled. Second, several Missouri districts participated in the Smarter Balanced field test that was also administered in the spring. Districts who participated in the Smarter Balanced field test may have experienced decreases in performance from 2013 to 2014 due to increased testing.

Lost Instructional Time

Almost all districts in the state reported missing some school due to snow. DESE provided preliminary data from 561 districts with the number of snow days taken by the district, the number of days the district made up following the snow days, the total number of school days originally planned by the district, and the total number of school days actually completed by the district..

Using preliminary data from 561 districts, Missouri districts had an average of 10.4 snow days (sd=5.52), 6.1 make-up days (sd=3.02), and 3.2 missed days (sd=6.41). Missed days was defined as the number of planned school days minus the number of completed school days.

To examine the possible effect of lost instructional time due to snow days, the following steps were taken:

1. School-level mean MAP scores and percentage at or above Proficient were computed for 2013 and 2014.
2. District information regarding snow days, make-up days, total planned days, and actual completed days was merged to the data. In some cases, school-level information regarding snow days, make-up days, total planned days, and actual completed days was available. For those cases, school-level information was used. For all other cases, district-level information was used.
3. Aggregate variables for snow days and missed days were created in order to examine patterns in the difference between the 2013 and 2014 test results.
4. Unweighted mean differences were examined at each grade level by aggregated snow days and by aggregated missed days.

These analyses were only completed for Communication Arts. If the analyses had uncovered evidence that lost instructional time was responsible for the changes in performance on MAP, then they would have been replicated on Mathematics.

Table 10 shows the average difference between 2013 and 2014 Communication Arts percentage at or above Proficient and mean Communication Arts MAP scores at each grade level by the number of snow days. Table 11 shows the average difference between 2013 and 2014 Communication Arts percentage at or above Proficient and mean Communication Arts MAP scores at each grade level by the number of school days missed.

If the number of snow days had an adverse impact, then it is expected that the schools with more snow days would have a larger decline in the percentage of students at or above Proficient and/or MAP mean SS than schools that had fewer snow days. Likewise, it is expected that schools that had more missed days would experience a larger decline in the percentage of students at or above Proficient and/or MAP mean SS than schools that had fewer missed days.

Table 10 and Table 11 do not confirm this expectation when disaggregated by grade. Schools that had more snow days do not routinely experience a larger decline than schools with fewer snow days. Likewise, schools with more missed school days do not routinely experience a larger decline in performance on MAP than schools with fewer missed school days. In both cases, the declines in percent at or above Proficient in 2014 compare to 2013 were of a similar magnitude regardless of snow days or missed days.

Table 10 Average Change in 2014 Percent At or Above Proficient and Average 2014 MAP Scale Score by Number of Snow Days, Communication Arts

			Difference in % at or above Proficient (2014-2013)		Difference in Average MAP SS (2014-2013)	
	Communication Arts	N	Mean	SD	Mean	SD
3	0 to 5 Snow Days	195	-6.18	14.70	-4.53	11.31
	6 to 10 Snow Days	602	-6.30	14.28	-5.84	11.54
	11 to 15 Snow Days	148	-5.56	13.09	-4.86	10.09
	16 to 20 Snow Days	93	-5.79	15.19	-5.54	10.52
	21 or More Snow Days	69	-5.41	15.95	-5.28	11.53
4	0 to 5 Snow Days	196	-7.13	14.52	-5.03	10.90
	6 to 10 Snow Days	600	-7.22	12.17	-5.04	10.50
	11 to 15 Snow Days	141	-6.91	14.28	-5.25	9.51
	16 to 20 Snow Days	93	-8.92	14.48	-5.91	10.37
	21 or More Snow Days	69	-8.81	16.75	-5.67	10.51
5	0 to 5 Snow Days	193	-2.19	13.96	-1.70	12.48
	6 to 10 Snow Days	567	-3.06	12.98	-1.95	10.21
	11 to 15 Snow Days	127	-4.31	12.39	-2.36	7.40
	16 to 20 Snow Days	81	-2.19	15.15	-1.07	10.00
	21 or More Snow Days	69	-2.93	16.31	-1.62	10.27
6	0 to 5 Snow Days	132	-3.41	18.44	-2.78	10.88
	6 to 10 Snow Days	364	-2.36	15.52	-1.79	10.74
	11 to 15 Snow Days	106	-1.60	12.58	-2.82	6.85
	16 to 20 Snow Days	79	-2.77	13.93	-0.19	9.02
	21 or More Snow Days	68	-2.67	16.24	-1.56	10.70
7	0 to 5 Snow Days	105	3.87	14.82	2.58	10.51
	6 to 10 Snow Days	310	0.37	14.77	0.55	12.12
	11 to 15 Snow Days	93	1.46	12.32	0.99	7.85
	16 to 20 Snow Days	80	-0.07	13.00	-0.27	8.95
	21 or More Snow Days	65	-1.25	13.36	-1.13	9.31
8	0 to 5 Snow Days	103	-4.27	16.12	-1.83	9.61
	6 to 10 Snow Days	307	-2.53	14.85	-1.66	8.54
	11 to 15 Snow Days	94	-5.50	12.73	-4.33	9.86
	16 to 20 Snow Days	79	-6.79	13.56	-4.04	8.61
	21 or More Snow Days	65	0.64	14.67	-1.27	9.29

Table 11 Average Change in 2014 Percent At or Above Proficient and Average 2014 MAP Scale Score by Number of Missed Days, Communication Arts

			Difference in % at or above Proficient (2014-2013)		Difference in Average MAP SS (2014-2013)	
	Communication Arts	N	Mean	SD	Mean	SD
3	Missed 0 to 5 Days	825	-6.20	14.46	-5.28	11.22
	Missed 6 to 10 Days	195	-6.06	13.36	-6.38	11.20
	Missed 11 to 15 Days	62	-5.42	15.37	-4.74	10.91
	Missed 16 to 20 Days	25	-3.74	16.89	-4.30	12.40
4	Missed 0 to 5 Days	819	-7.05	13.15	-5.04	10.68
	Missed 6 to 10 Days	192	-8.45	12.73	-5.63	9.25
	Missed 11 to 15 Days	63	-8.11	16.57	-6.36	10.01
	Missed 16 to 20 Days	25	-9.57	17.48	-3.34	11.72
5	Missed 0 to 5 Days	768	-2.82	13.42	-1.83	10.77
	Missed 6 to 10 Days	184	-3.57	12.06	-2.12	8.32
	Missed 11 to 15 Days	60	-4.58	17.33	-2.42	10.49
	Missed 16 to 20 Days	25	0.47	15.66	0.28	10.66
6	Missed 0 to 5 Days	507	-2.83	16.51	-2.18	10.72
	Missed 6 to 10 Days	160	-1.67	12.10	-1.11	7.82
	Missed 11 to 15 Days	58	-3.73	16.87	-3.22	9.95
	Missed 16 to 20 Days	24	1.57	12.50	1.26	10.91
7	Missed 0 to 5 Days	437	1.54	14.80	1.36	11.55
	Missed 6 to 10 Days	136	0.54	12.38	-0.33	7.88
	Missed 11 to 15 Days	58	0.07	13.01	-0.24	10.74
	Missed 16 to 20 Days	22	-8.20	10.97	-4.26	6.93
8	Missed 0 to 5 Days	436	-3.09	15.37	-1.90	9.42
	Missed 6 to 10 Days	133	-5.85	12.17	-4.06	7.64
	Missed 11 to 15 Days	57	-1.86	14.90	-1.48	9.11
	Missed 16 to 20 Days	22	0.26	13.51	-2.54	8.40

Smarter Balanced Field Test Participation

Several schools participated in the Smarter Balanced field test. The test results from these schools were compared with the test results from those schools that did not participate in the Smarter Balanced field test. Table 12 reports the average difference in percent at or above Proficient for Communication Arts between 2013 and 2014 as well as the average difference in mean Communication Arts MAP scale score between 2013 and 2014 disaggregated by Smarter Balanced field test participation. Table 12 does not show a clear pattern in the results when disaggregated by field test participation. At some grade levels, schools that participated in the field test show smaller decreases in performance from the previous year compared to non-participating schools. In other grade levels, the

opposite was observed: participating schools had larger decrease in MAP performance than non-participating schools.

These analyses were only completed for Communication Arts. If the analyses had uncovered evidence that participation in the Smarter Balanced field test was responsible for the changes in performance on MAP, then they would have been replicated on Mathematics.

Table 12 Average Change in 2014 Percent At or Above Proficient and Average 2014 MAP Scale Score by Smarter Balanced Field Test Participation, Communication Arts

Grade	Participated in SBAC	N	Difference in % at or above Proficient (2014-2013) Mean	SD	Difference in Average MAP SS (2014-2013) Mean	SD
3	No	1071	-6.22	14.52	-5.60	11.58
	Yes	38	-3.42	10.70	-2.22	8.33
4	No	1053	-7.31	13.56	-5.04	10.76
	Yes	48	-8.86	9.80	-6.58	7.84
5	No	1006	-2.99	13.57	-1.85	10.55
	Yes	33	-1.67	12.77	-0.57	9.22
6	No	729	-2.56	15.98	-1.93	10.29
	Yes	22	-5.06	9.46	-3.29	4.68
7	No	626	0.76	14.15	0.69	10.61
	Yes	29	4.39	14.27	0.88	13.17
8	No	610	-3.41	14.87	-2.35	9.15
	Yes	40	-3.30	11.81	-1.52	7.24

Summary of Test Administration Comparability

There were differences between the administration conditions of the 2014 test compared to previous years. As discussed above, schools missed more of their school year than usual due to snow. Also, several schools administered the Smarter Balanced field test. The analyses in this section did not reveal a pattern due to missed school days or snow days. Nor did the analyses of Smarter Balanced participation show a clear pattern where schools that participated in the Smarter Balanced field test performed better or worse than schools that did not participate. Given the analyses in this section, there is no reason to believe that changes in administration conditions were the cause of the decline in test scores.

Comparability of Test Construct

Next, the comparability of the test construct was explored. It is expected that test blueprint is the same between test administrations. It is expected that the distribution of the items will be similar (+/-10 percentage points) across the content standards in both 2013 and 2014.

For this analysis, the distribution of items by content standard in 2014 was compared to the distribution of items by content standard in 2013. Table 13, Table 14, and Table 15 show the difference between 2014 and 2013 in the percent of items measuring each content standard for Communication Arts, Mathematics, and Science, respectively.

Each table shows that the distribution of items by content standard is similar between the two years. Most of the differences in the distribution of items are within five percentage points when 2014 is compared to 2013. The construct being measured appears to be similar between the two forms.

Table 13 Change in Percentage of Items Measuring Each Content Standard in 2014 compared to 2013, Communication Arts

	Change in Percentage of Items Measuring Each Standard (2014 form – 2013 form)					
Content Standard	3	4	5	6	7	8
Speaking/Writing Standard English	0.7%	-5.1%	0.7%	-1.1%	1.8%	0.0%
Reading—Fiction & Nonfiction	-1.6%	4.8%	-0.9%	0.9%	-2.4%	-0.1%
Writing Formally & Informally	0.9%	0.4%	0.2%	0.2%	0.7%	0.1%

Table 14 Change in Percentage of Items Measuring Each Content Standard in 2014 compared to 2013, Mathematics

	Change in Percentage of Items Measuring Each Standard (2014 form – 2013 form)					
Content Standard	3	4	5	6	7	8
Number and Operations	4.2%	-2.7%	0.8%	2.9%	3.6%	1.7%
Algebraic Relationships	-3.2%	0.4%	0.6%	-5.1%	0.3%	-0.4%
Geometric and Spatial Relationships	-2.8%	-0.5%	-2.8%	1.4%	-4.4%	-6.3%
Measurement	0.6%	0.8%	0.6%	-0.2%	0.2%	2.2%
Data and Probability	1.1%	2.1%	0.9%	1.2%	0.3%	2.8%

Table 15 Change in Percentage of Items Measuring Each Content Standard in 2014 compared to 2013, Science

Content Standard	Change in Percentage of Items Measuring Each Standard (2014 form – 2013 form) 5	8
Matter and Energy	3.0%	1.3%
Force and Motion	-3.4%	-2.3%
Living Organisms	-2.2%	1.3%
Ecology	-2.1%	-1.1%
Earth Systems	4.2%	1.3%
Universe	-0.8%	3.7%
Scientific Inquiry	-1.5%	-3.2%
Science, Technology, and Human Activity	2.8%	-1.1%

Summary of Test Construct Comparability

The test construct measured in 2014 appears to be comparable to the test construct measured in 2013.

Comparability of Psychometric Characteristics of the Total Test

The next set of analyses examined the psychometric characteristics of the total test. In particular, DESE asked if this set of forms was more difficult than past forms. This set of analyses examined overall test statistics, form difficulty, as well as the item calibration.

Overall Test Statistics

Table 16, Table 17, and Table 18 show the total number of items, points, mean p-value, and test reliability in 2013 and 2014 for Communication Arts, Mathematics, and Science, respectively.

The test reliabilities (as measured by Cronbach's Alpha) are at or above 0.90 for all grades and content areas, except Grade 8 Mathematics. The test reliability for Grade 8 Mathematics is 0.89. All test reliabilities are acceptable for a state-level summative test.

The mean p-value is often used as an indication of the difficulty of the form; however, Missouri forms are calibrated and scaled using item response theory. The use of IRT allows us to take into account the relative difficulty of the items in one form compared to another form. Using IRT, a more precise measure of form difficulty are the test characteristic curves presented in the next session. The p-values in Table 16, Table 17, and Table 18 provide an indication of how well the 2014 students did on the form they were administered compared to the 2013

students. It shows the average proportion of items that students answered correctly.

Table 16 shows that there tends to be a small decline in 2014 mean p-value from 2013 in all grades within Communication Arts. The largest decline (-0.04) in p-value occurs in Grade 3 Communication Arts, where the mean p-value was 0.76 in 2013 and is 0.72 in 2014. This means that the students were answering a lower number of items correctly in 2014 compared to 2013.

Table 17 shows that there are small increases in 2014 mean p-value from 2013 in Grade 4 and 8 Mathematics and small declines in Grades 3, 6, and 7. Table 18 shows that the mean p-values in Science tend to be similar between the two years.

One notable difference observed in Table 16 and Table 17 is that the Communication Arts and Mathematics tests tended to be shorter in 2014 than they were in 2013. For example, the Grade 4 Mathematics scale score was based on 51 items in 2014 and 62 items in 2013.

Table 16 Overall 2013 and 2014 Test Statistics (Total Number of Items, Total Number of Points, Mean P-Value, and Reliability), Communication Arts

	2013				2014			
Grade	**Total Items**	**Total Points**	**Mean *p*-value (SD)**	**Reliability**	**Total Items**	**Total Points**	**Mean *p*-value (SD)**	**Reliability**
3	57	65	0.76 (0.14)	0.91	52	59	0.72 (0.15)	0.91
4	56	61	0.77 (0.15)	0.91	51	55	0.76 (0.15)	0.90
5	57	61	0.74 (0.16)	0.91	52	54	0.72 (0.13)	0.91
6	56	60	0.72 (0.15)	0.91	49	53	0.72 (0.14)	0.90
7	63	70	0.71 (0.16)	0.91	57	65	0.70 (0.16)	0.91
8	60	64	0.70 (0.16)	0.91	56	60	0.69 (0.19)	0.90

Table 17 Overall 2013 and 2014 Test Statistics (Total Number of Items, Total Number of Points, Mean P-Value, and Reliability), Mathematics

	2013				2014			
Grade	Total Items	Total Points	Mean *p*-value (SD)	Reliability	Total Items	Total Points	Mean *p*-value (SD)	Reliability
3	55	59	0.82 (0.13)	0.91	53	57	0.80 (0.12)	0.91
4	62	69	0.75 (0.13)	0.92	51	57	0.77 (0.14)	0.91
5	57	61	0.73 (0.15)	0.92	55	59	0.73 (0.14)	0.92
6	58	62	0.71 (0.16)	0.91	59	63	0.70 (0.14)	0.91
7	61	65	0.66 (0.14)	0.92	60	64	0.65 (0.17)	0.92
8	61	68	0.56 (0.18)	0.90	57	64	0.57 (0.21)	0.89

Table 18 Overall 2013 and 2014 Test Statistics (Total Number of Items, Total Number of Points, Mean P-Value, and Reliability), Science

	2013				2014			
Grade	Total Items	Total Points	Mean *p*-value (SD)	Reliability	Total Items	Total Points	Mean *p*-value (SD)	Reliability
5	64	82	0.64 (0.20)	0.91	64	79	0.64 (0.20)	0.90
8	65	85	0.62 (0.22)	0.92	66	84	0.63 (0.16)	0.92

Form Difficulty

Test developers construct test forms to be of equivalent difficulty from year to year. It is not expected that the test forms will have the same exact difficulty from year to year. DESE provided test characteristic curves (TCCs) for each grade for all testing years (2006 to 2014) in order to look at the relative difficulty of all forms in relationship to each other. Figure 1through Figure 14 show the TCCs from 2006 to 2014 for all grades and content areas.

The TCCs were investigated to see if the 2014 forms were routinely more difficult or easier than the 2013 forms. This analysis was of particular importance in Grades 3 and 4 Communication Arts and Grade 4 Mathematics where the largest drop in MAP performance was observed.

Examination of the TCCs does not reveal a particular pattern where the 2014 was always harder or always easier than the other years. All 2014 TCCs appear to

represent tests of similar difficulty to previous years. In the three grade/content area of particular interest:

- The 2014 Grade 3 Communication Arts TCC appears to be in the center of the TCCs from 2006 to 2013, indicating it was easier than some past forms and more difficult than others.
- The 2014 Grade 4 Communication Arts TCC appears to be in the center of the 2006 to 2013 TCCs for the lower half of the test scale. For the upper half, the 2014 TCC appears to be one of the easier forms.
- The 2014 Grade 4 Mathematics TCC appears to be one of the easier forms, if not the easiest form, compared to the TCCs from 2006 to 2013.

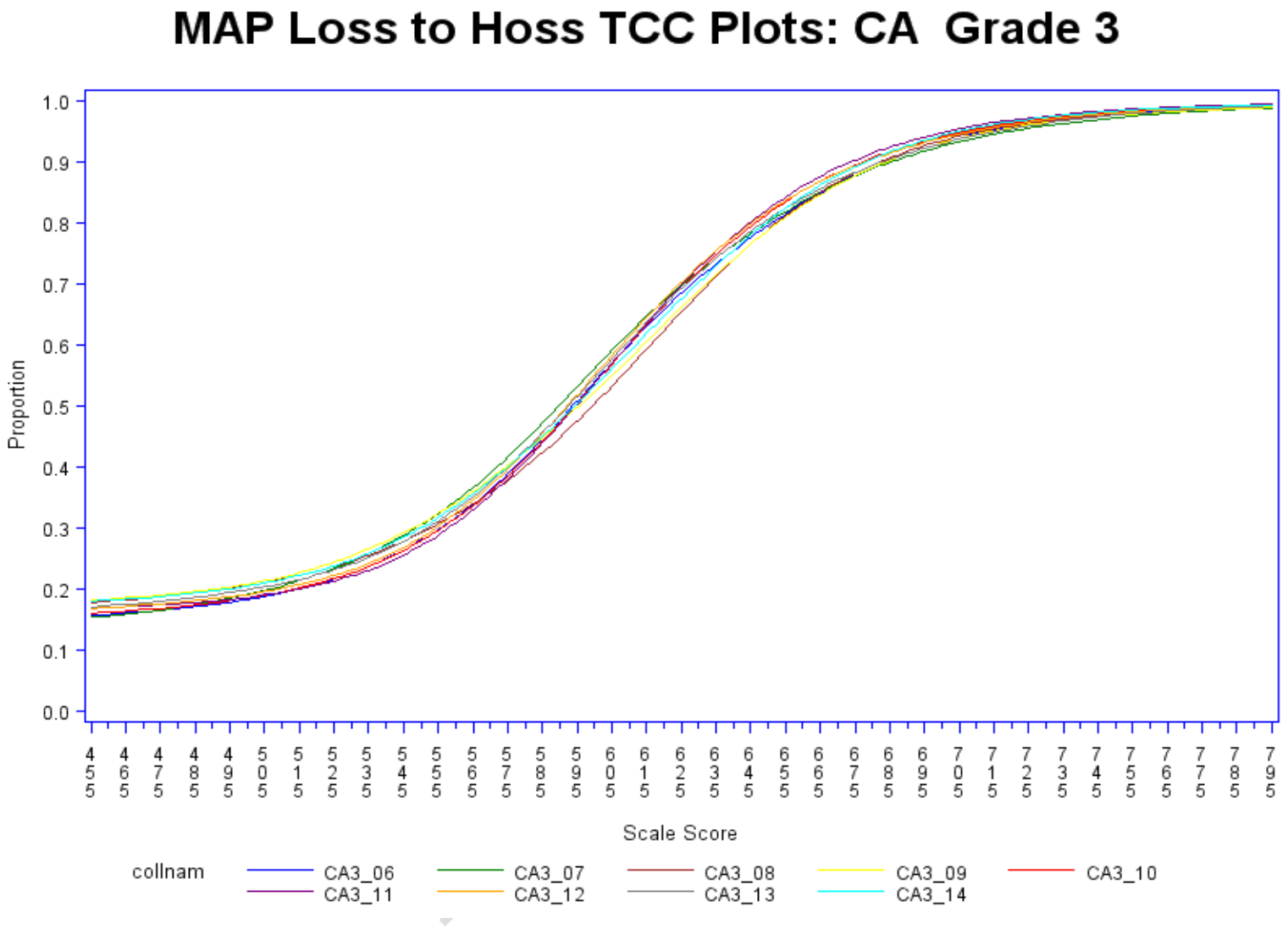

Figure 1 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 3 Communication Arts

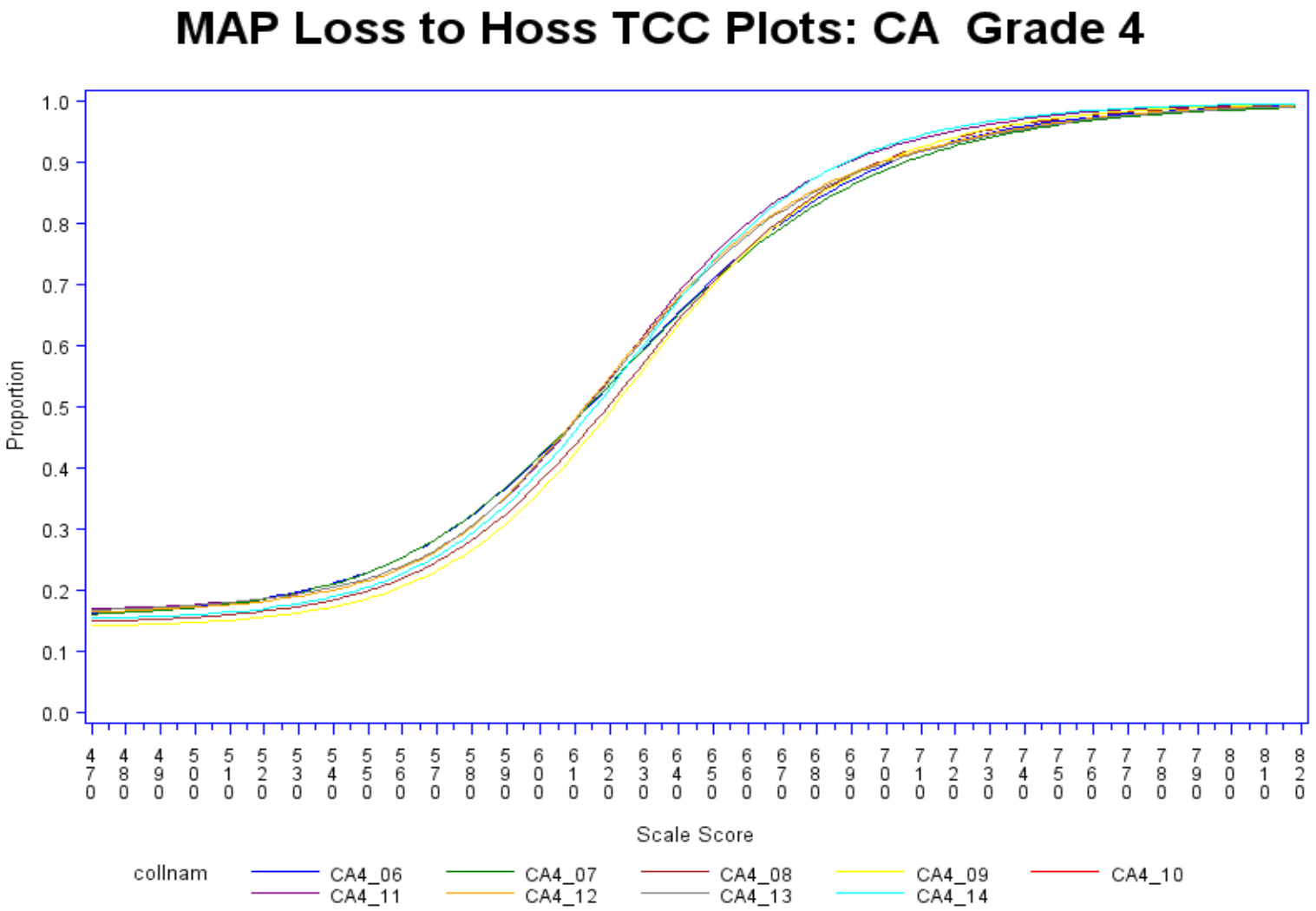

Figure 2 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 4 Communication Arts

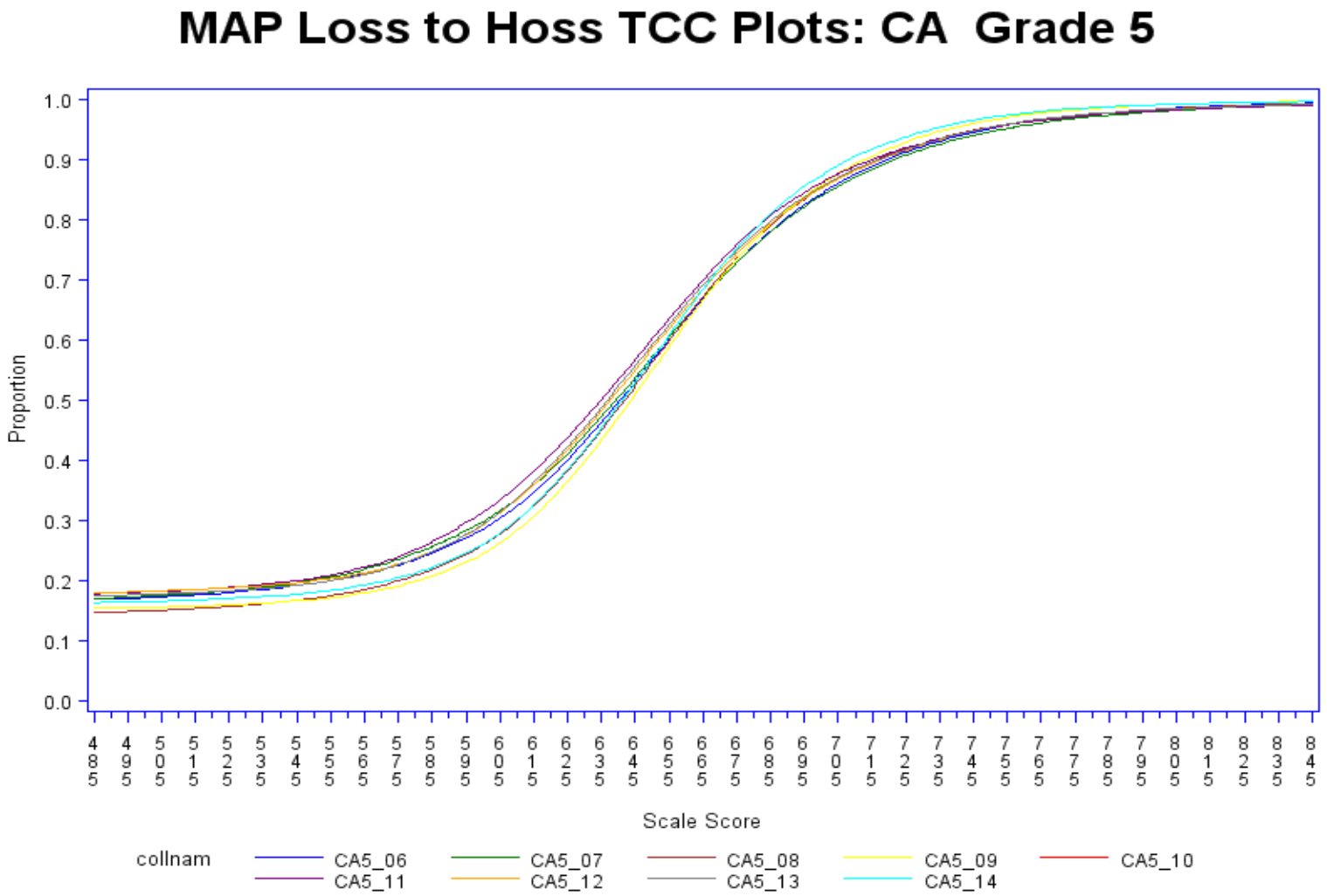

Figure 3 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 5 Communication Arts

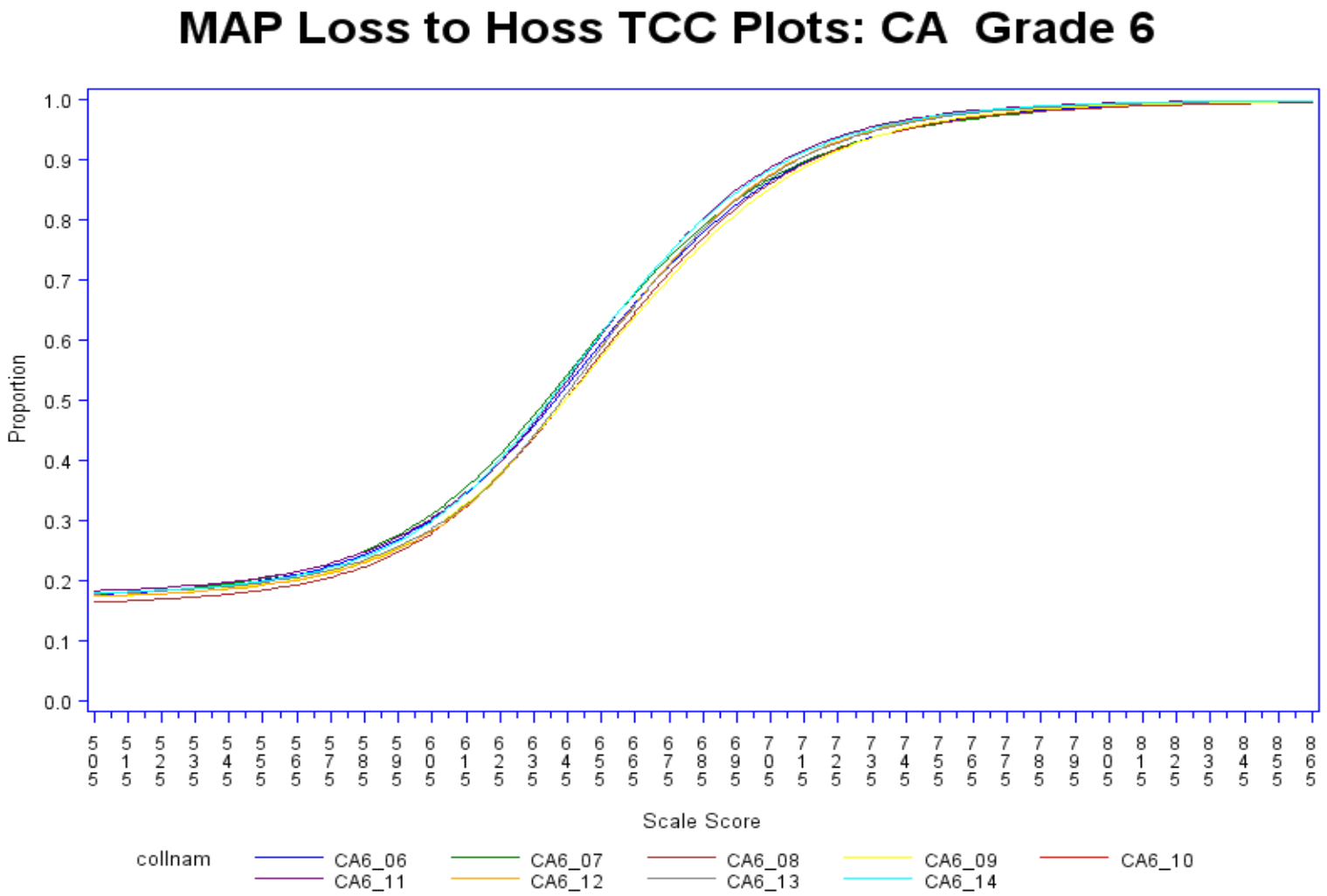

Figure 4 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 6 Communication Arts

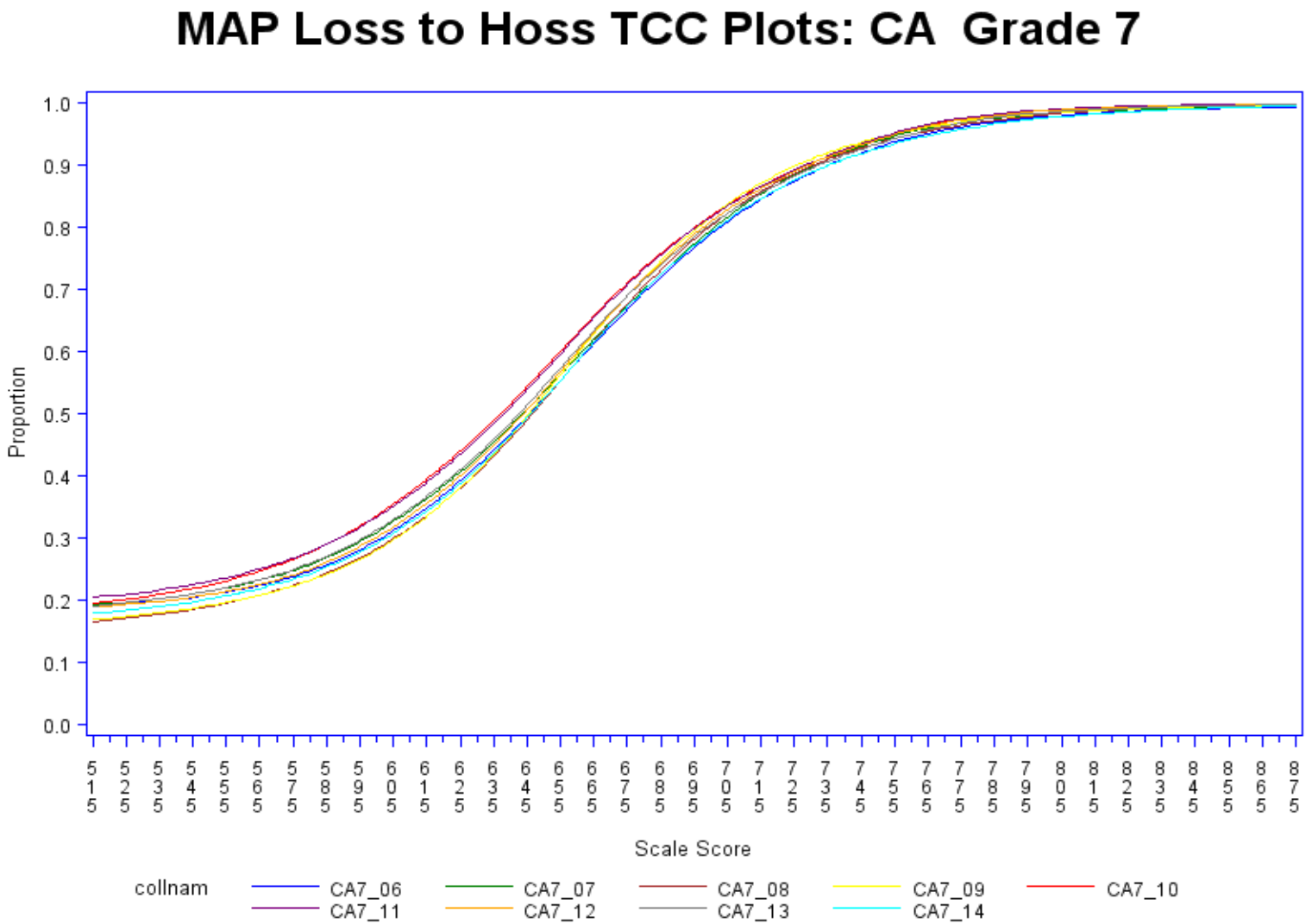

Figure 5 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 7 Communication Arts

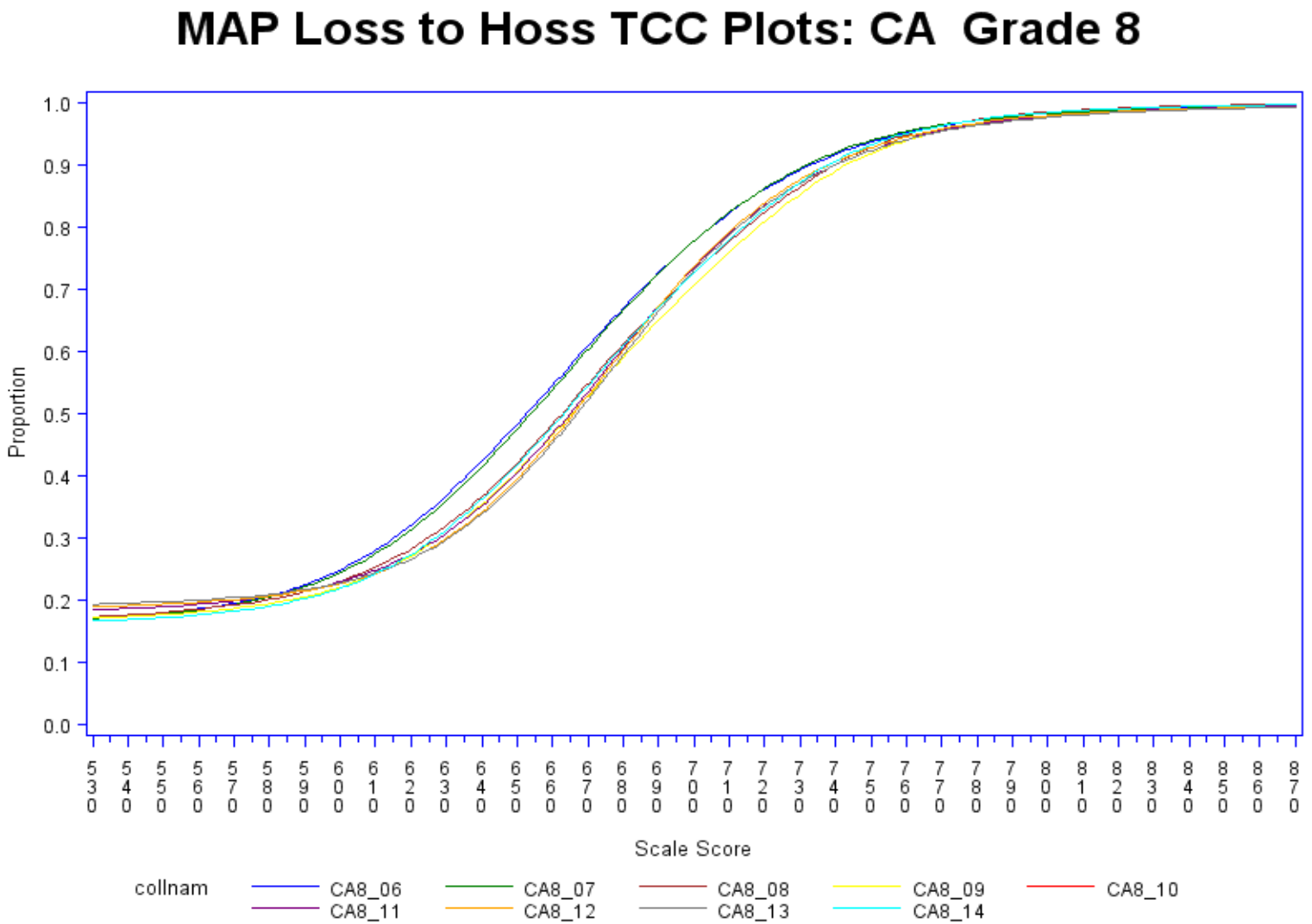

Figure 6 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 8 Communication Arts

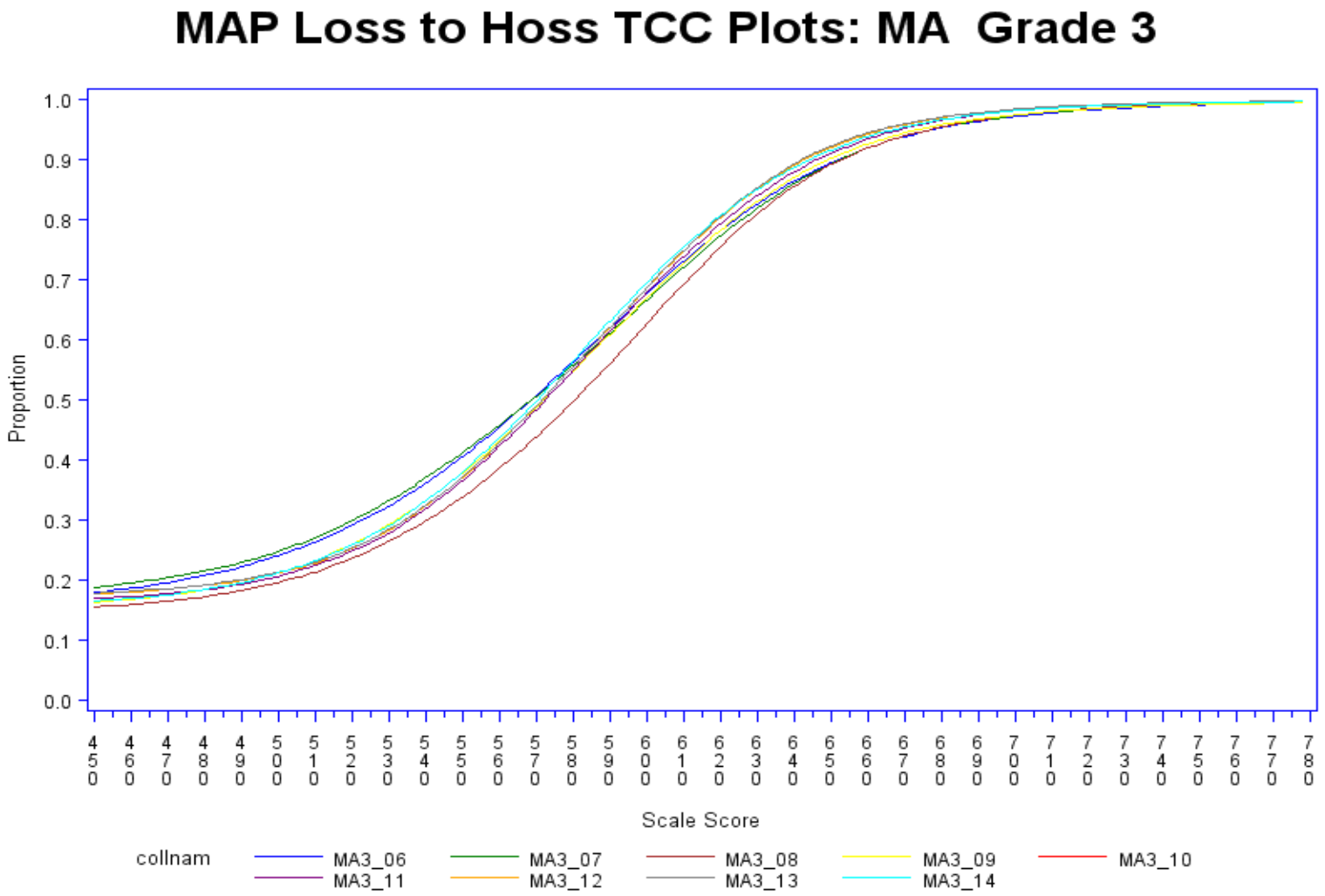

Figure 7 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 3 Mathematics

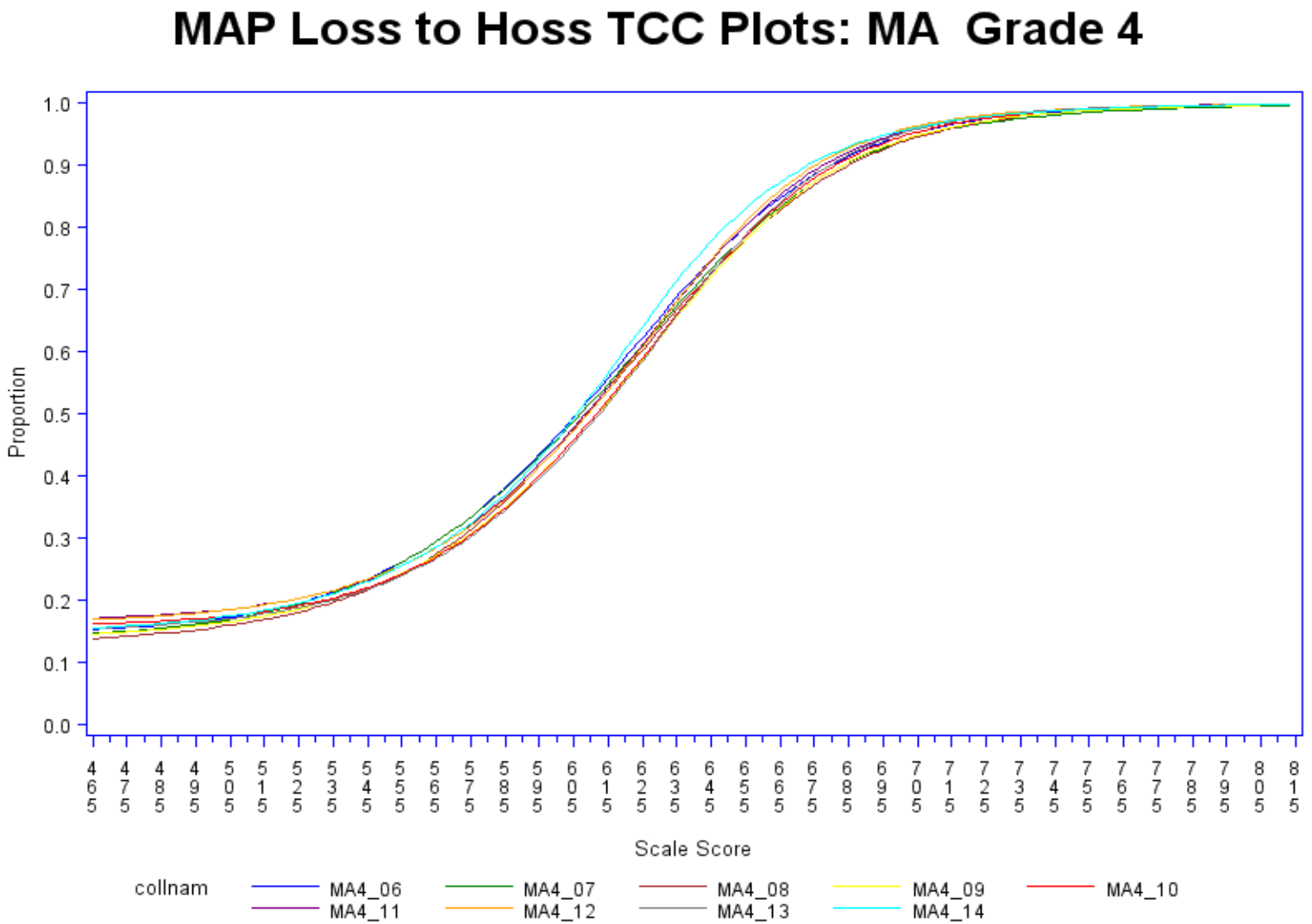

Figure 8 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 4 Mathematics

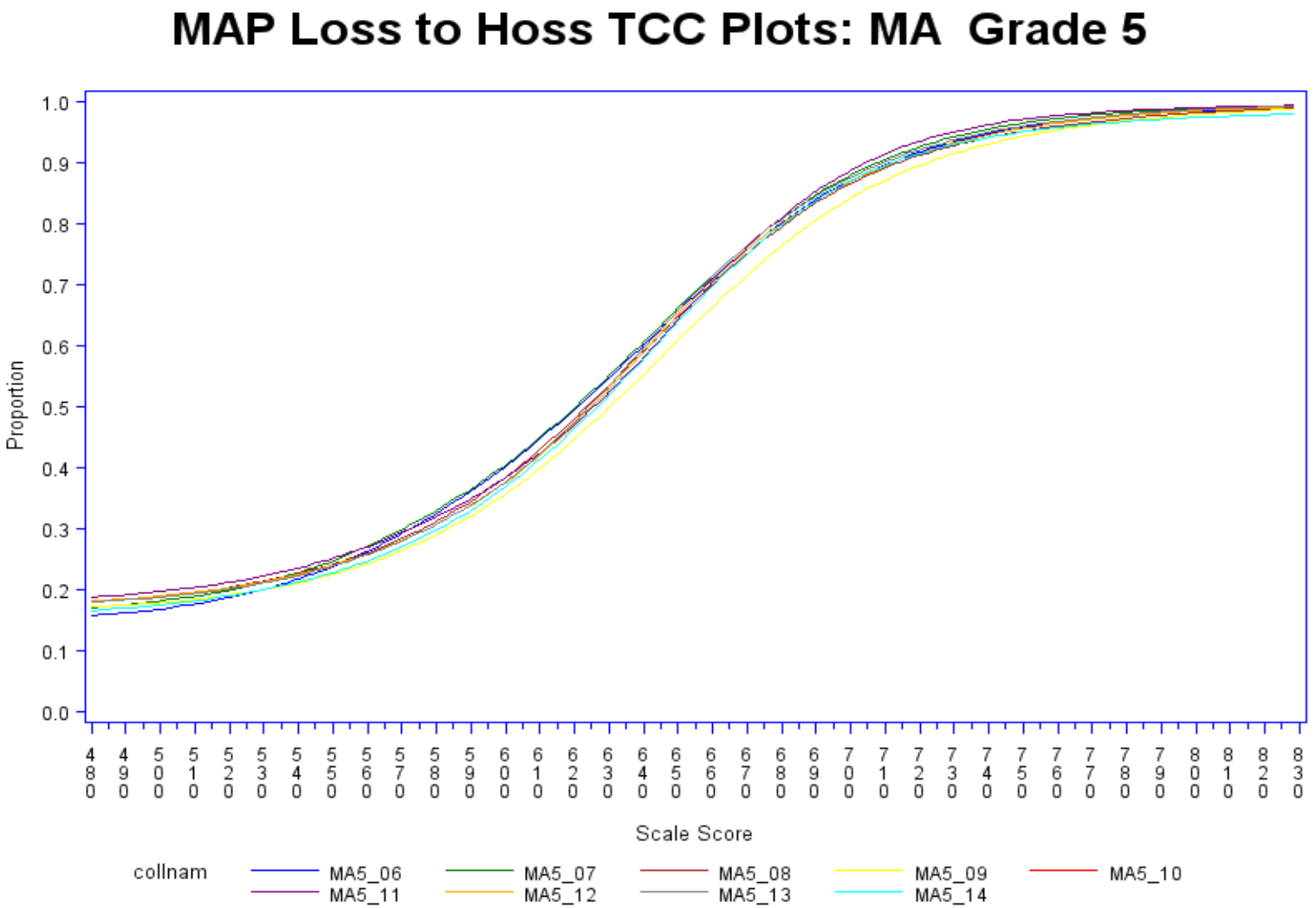

Figure 9 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 5 Mathematics

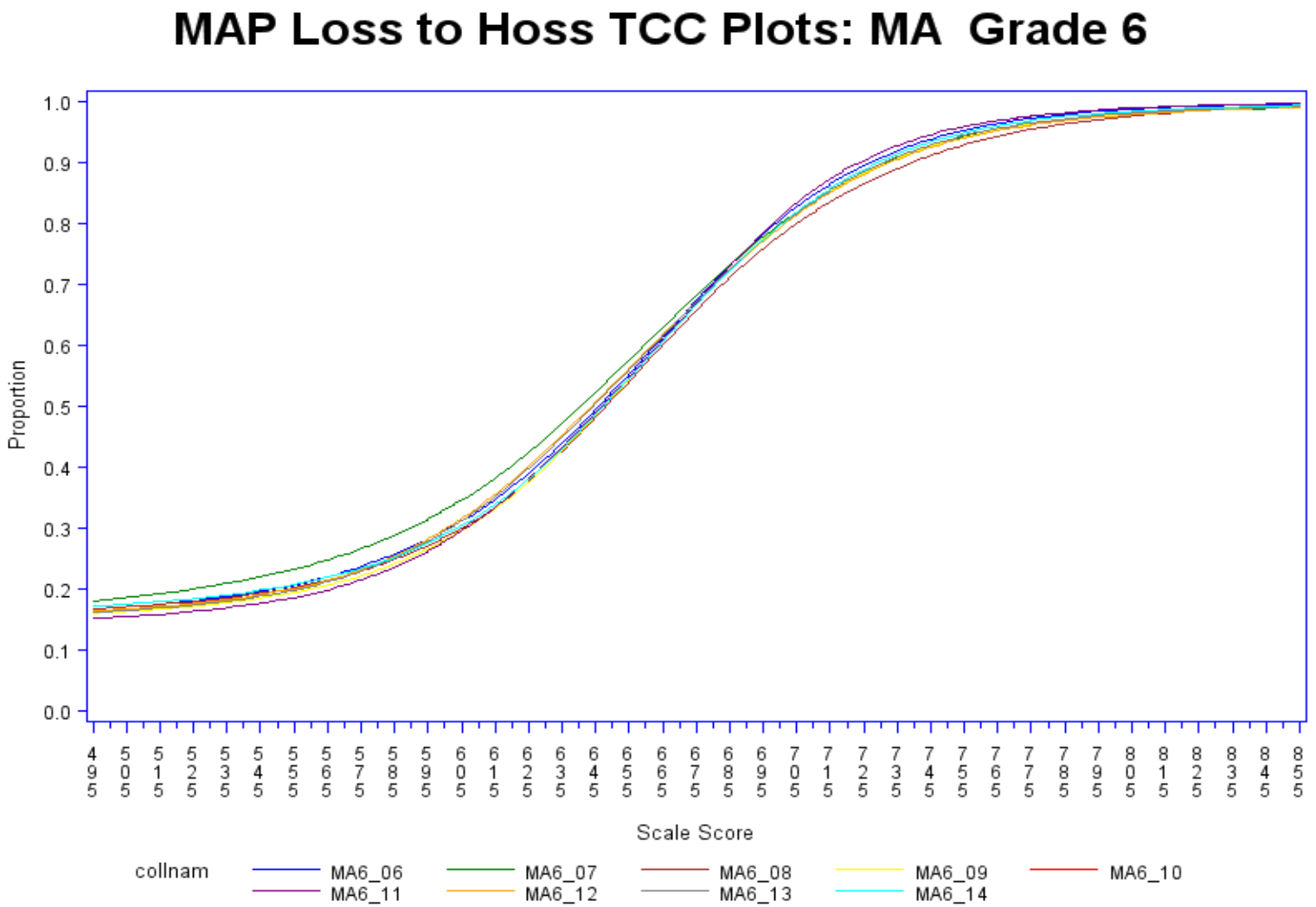

Figure 10 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 6 Mathematics

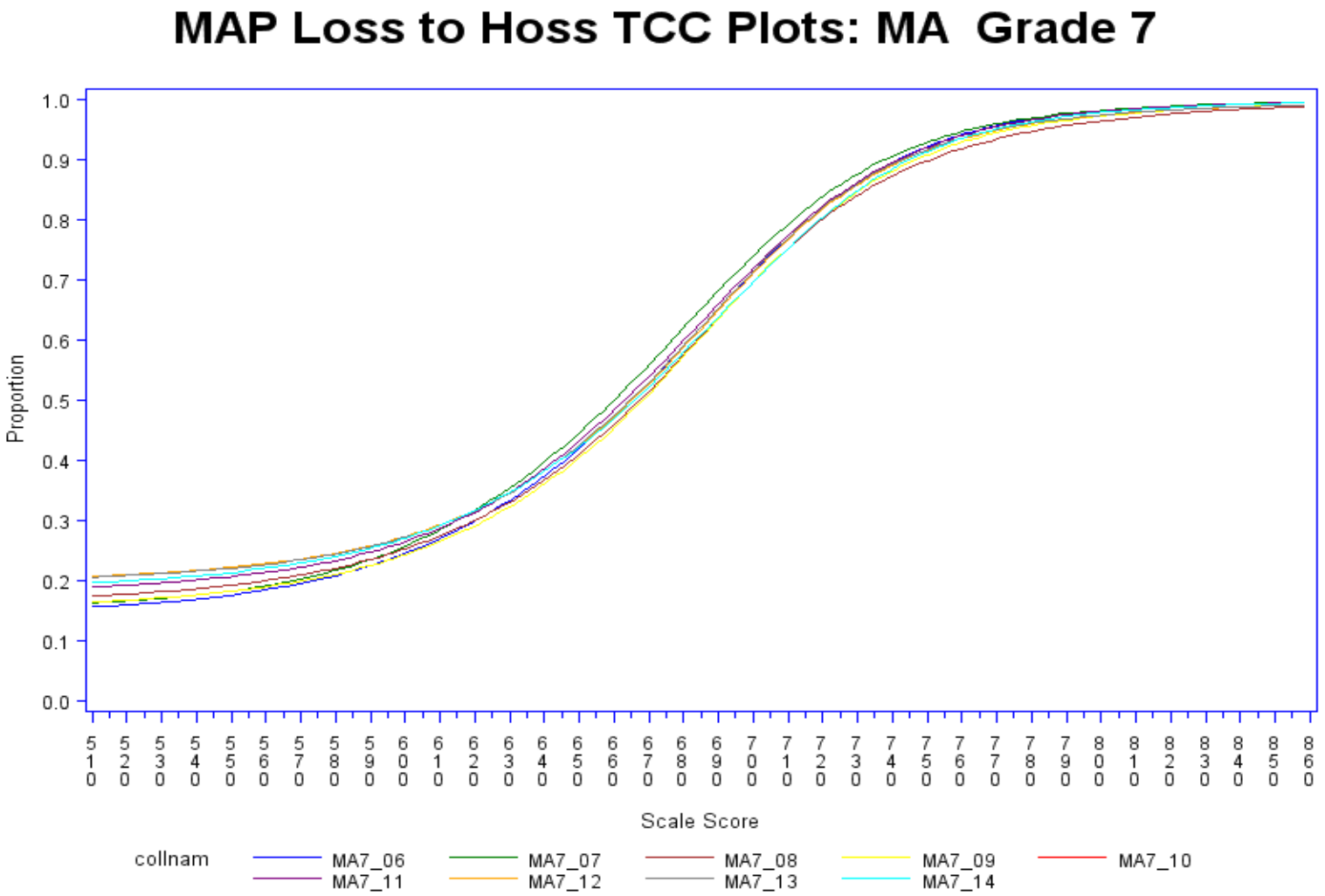

Figure 11 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 7 Mathematics

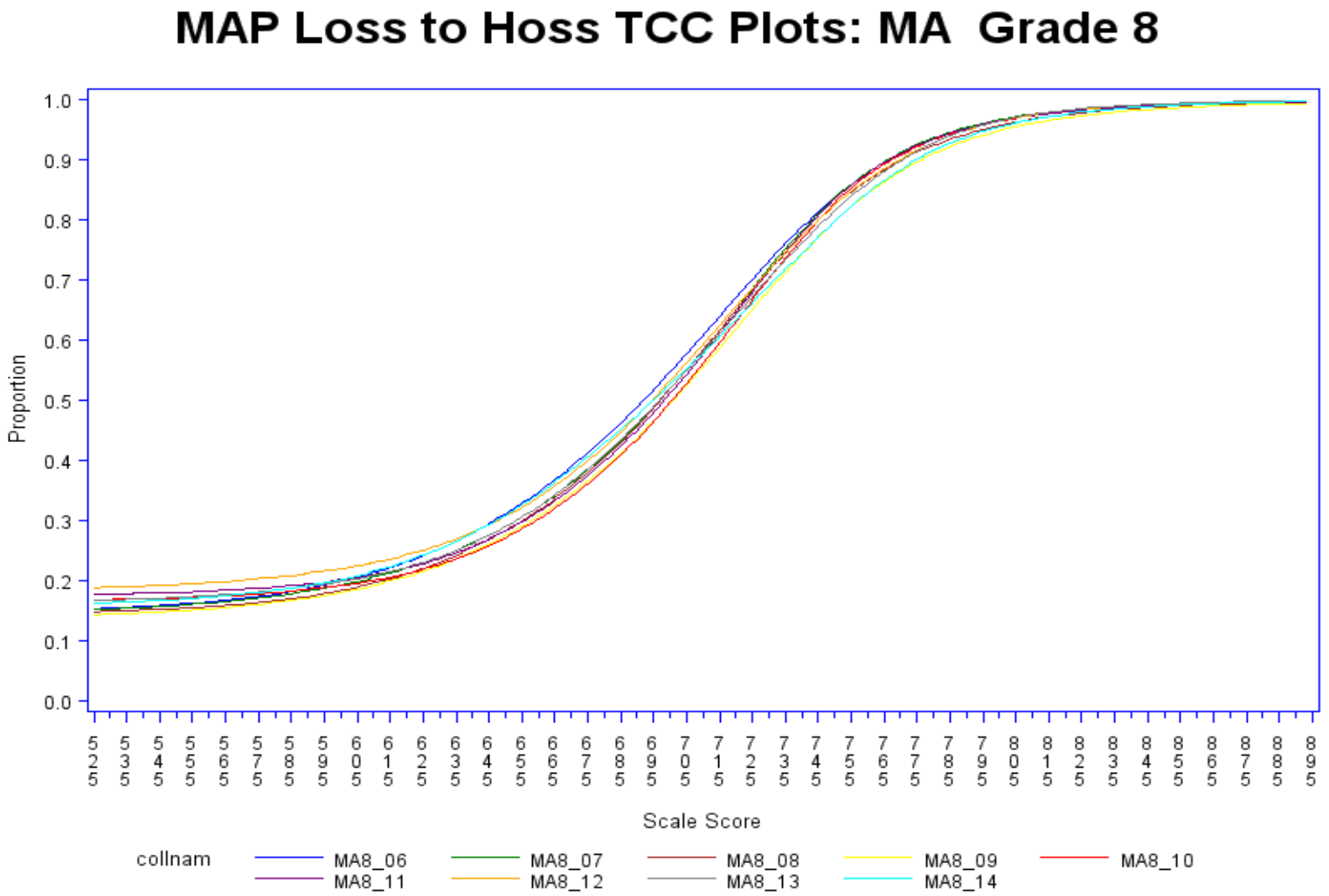

Figure 12 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 8 Mathematics

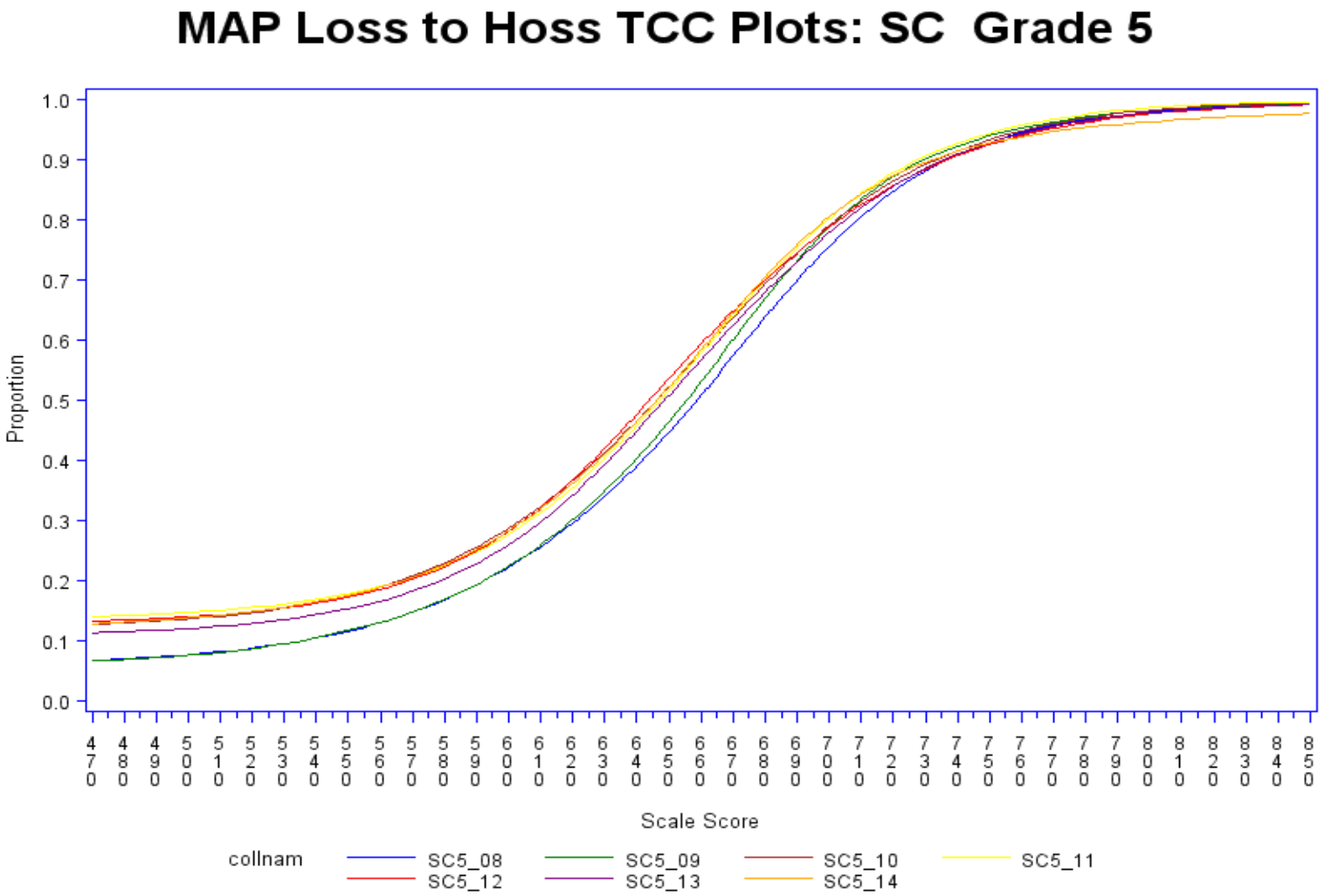

Figure 13 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 5 Science

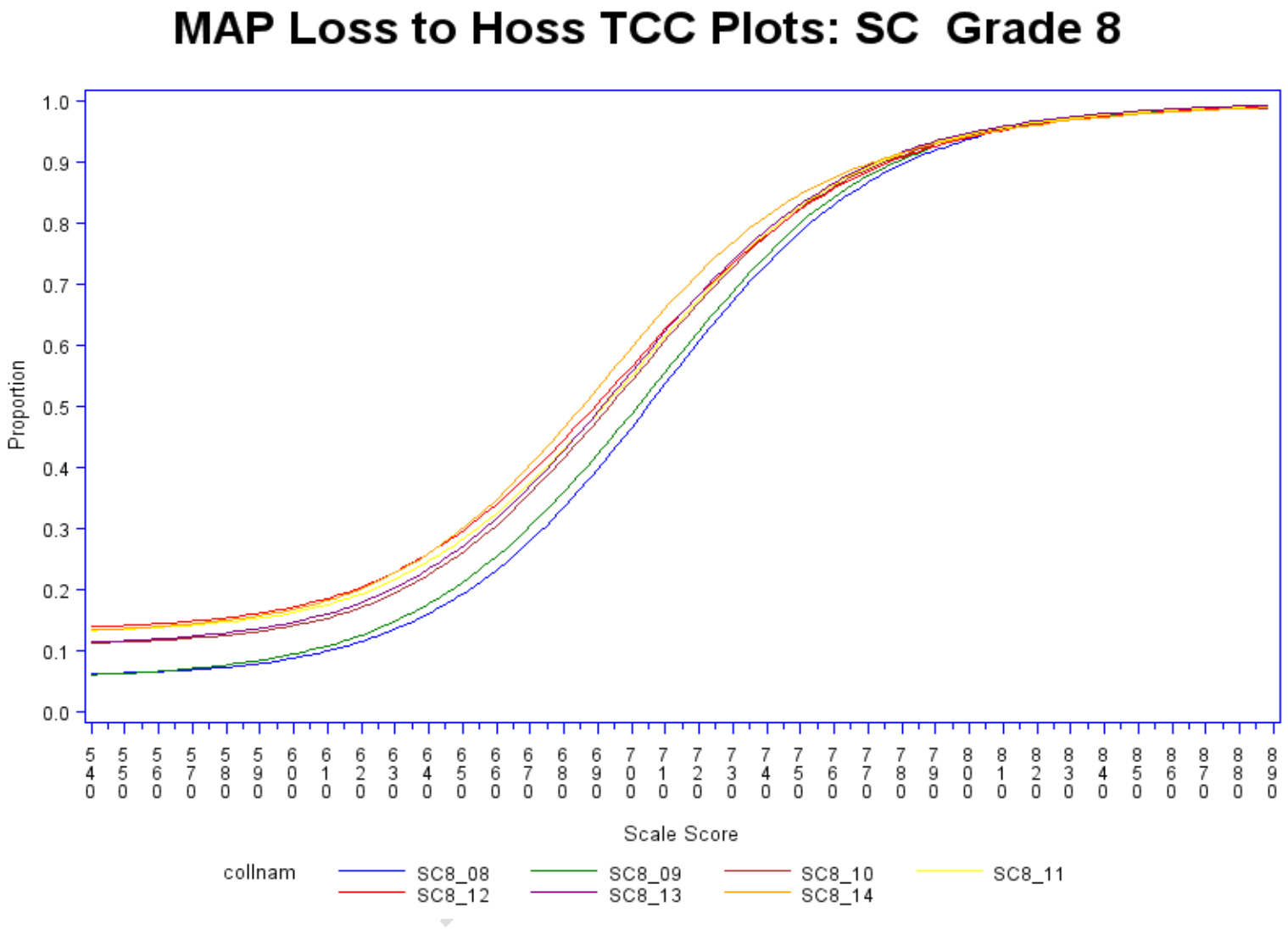

Figure 14 MAP Test Characteristic Curves for 2006 through 2014 Test Forms, Grade 8 Science

Item Calibration

DESE provided CTB's output from item calibrations. This was examined to see if a different type of calibration sample was used in 2014 compared to previous years. The 2013 MAP Technical Report states that the 2013 calibration sample comprised "at least 80% of the student population for that grade" (p. 67). Tables 6.2 through 6.4 in the 2013 Technical Report show that almost the entire population was used for calibrating items in 2013.

Table 19, Table 20, and Table 21 show the size of the calibration sample and the number of items with that demonstrated poor fit with the IRT model in both 2014 and 2013 for Communication Arts, Mathematics, and Science, respectively. These tables show that nearly the entire census population was used in the calibration sample in both years. These tables also show that there were very few misfitting items in both testing years.

Table 19 Number of Students in 2013 and 2014 Calibration Samples and Total Number of Misfit Items in 2013 and 2014, Communication Arts

Grade	2014		2013	
	Calibration Sample	# of misfit items	Calibration Sample	# of misfit items 2013
3	66871	0	66479	1
4	66365	1	65849	1
5	65885	0	65689	1
6	65747	0	66373	0
7	66580	0	67041	0
8	66813	1	64150	1

Table 20 Number of Students in 2013 and 2014 Calibration Samples and Total Number of Misfit Items in 2013 and 2014, Mathematics

Grade	2014		20133	
	Calibration Sample	Number of Misfit Items	Calibration Sample	Number of Misfit Items
3	67047	1	66598	1
4	66517	1	65980	0
5	66008	1	65835	1
6	65803	0	66451	1
7	65811	0	66277	1
8	51957	1	49988	0

Table 21 Number of Students in 2013 and 2014 Calibration Samples and Total Number of Misfit Items in 2013 and 2014, Science

Grade	2014		20133	
	Calibration Sample	Number of Misfit Items	Calibration Sample	Number of Misfit Items
5	65893	0	65814	2
8	66903	0	64210	2

Summary of Psychometric Characteristics

The psychometric characteristics of the 2014 assessment are similar to those of the 2013 assessment. Classical measures of test reliability tend to be similar in both years of testing. The mean p-values indicate that students performed about as well in 2014 as they did in 2013. The TCCs from the IRT analyses show that the tests tend to be in the same range of difficulty as previous forms. One notable difference is that the 2014 tests tend to be shorter than the 2013 form.

Comparability of the Internal Anchor

The final set of analyses examined the comparability of the internal anchor set across years. This set of analyses examined the construction of the anchor, the construct of the anchor, the placement of the anchors, the results of CTB's Stocking & Lord equating, and the results of CTB's anchor evaluation.

DESE provided CTB's output from the construction of the internal anchor. Specifications were not provided; however, it can be deduced from the output provided by CTB that there were blueprint requirements for the anchor, a minimum length for the anchor, and TCC requirements for the anchor.

Length of the Internal Anchor

Kolen and Brennan (2004) recommend that anchors are, at least, 20% of the total test length for test containing 40 or more items (p.271). Table 22 shows the number of anchor items and each anchor's percentage of total test length. The anchor set is between 20% to 25% of the total test length. These anchors are within accepted industry standards.

In Grade 5 Mathematics, Grade 5 Science, and Grade 8 Science, the items listed in the original anchor selection were not the items used in the final anchor. In Grade 5 Mathematics, two items were removed and one was added. In Grade 5 Science[1], the original anchor selection consisted of 17 items and 22 items were used in the final

[1] The items identified as anchors in Grade 5 Science was different in the metadata file and CTB's Stocking & Lord output. For this paper, we used the items identified as anchors in CTB's Stocking & Lord output.

anchor. In Grade 8 Science, the original anchor consisted of 17 items and 25 items were used in the final anchor. The reason for these changes are unknown.

Table 22 Number of Anchor Items, Total Number of Test Items, and Anchors as a Percent of Total test by Grade Level and Content Area

Content Area	Grade	Number of Anchors	Total Number of Items	Percent of Total Test
	3	12	52	23.1%
	4	12	51	23.5%
CA	5	12	52	23.1%
	6	13	49	26.5%
	7	14	57	24.6%
	8	13	56	23.2%
	3	13	53	24.5%
	4	14	51	27.5%
MA	5*	11	55	20.0%
	6	14	59	23.7%
	7	14	60	23.3%
	8	13	57	22.8%
SC	5*	22	64	34.4%
	8*	25	66	37.9%

*Original anchor selection not provided.

The internal anchor appears to have been selected from an item pool consisting of items from multiple MAP forms administered from 2006 to 2013. Table 23 delineates the year in which each anchor item was originally administered. For Communication Arts and Science at all grade levels, the majority of the items were pulled from the 2009 administration. For Grades 3, 4, and 5 Mathematics, the majority of the items were pulled from the 2009 administration. For Grades 6, 7, and 8 Mathematics, the majority of the items were pulled from the test forms administered in 2010 and 2011.

There are two items for which the previous administration could not be identified. Item metadata files were only provided for 2009 to 2013. It is assumed that the two items where no year was identified came from 2006 to 2008.

Table 23 Original Testing Year of Anchor Items

Content Area	Grade	2009	2010 & 2011	2012 & 2013	No Year	Grand Total
Communication Arts	3	11		1		12
	4	9		2	1	12
	5	11		1		12
	6	10		2	1	13
	7	12		2		14
	8	11		2		13
Math	3	9	3	1		13
	4	10	2	2		14
	5	6	2	3		11
	6	2	10	1	1	14
	7		14			14
	8	3	10			13
Science	5	15	7			22
	8	25				25

The 2014 TCC for the internal anchor was selected to match the 2012 anchor TCC (2014, internal communication between CTB and DESE). The test forms and anchors were constructed in May 2013 before 2013 test data was available.

DESE provided CTB's anchor selection TCCs. Figure 15 through Figure 25 show the TCC for the selected anchor compared to the 2012 anchor TCC. Grade 5 Mathematics and Grades 5 and 8 Science are not included in these figures since these anchor sets changed from the original anchor sets.

The Y-axis in these figures shows the expected proportion correct and the x-axis shows the scale score range. CTB selected the items for the 2014 anchor to minimize the difference between the 2012 TCC and the 2014 TCC at each point along the scale score range. Table 24 shows the maximum difference in proportion correct between the 2012 and 2014 anchor TCCs. All differences are within five percentage points.

Table 24 Absolute Maximum Difference in Proportion Correct between 2012 and 2014 Anchor TCCs

Content Area	Grade	Absolute Maximum Difference
CA	3	0.040
	4	0.015
	5	0.014
	6	0.008
	7	0.005
	8	0.004
MA	3	0.032
	4	0.008
	5*	
	6	0.010
	7	0.005
	8	0.026
SC	5*	
	8*	

*Not available

MO Anchor Grade 3 CA TCC Plot from Form Construction

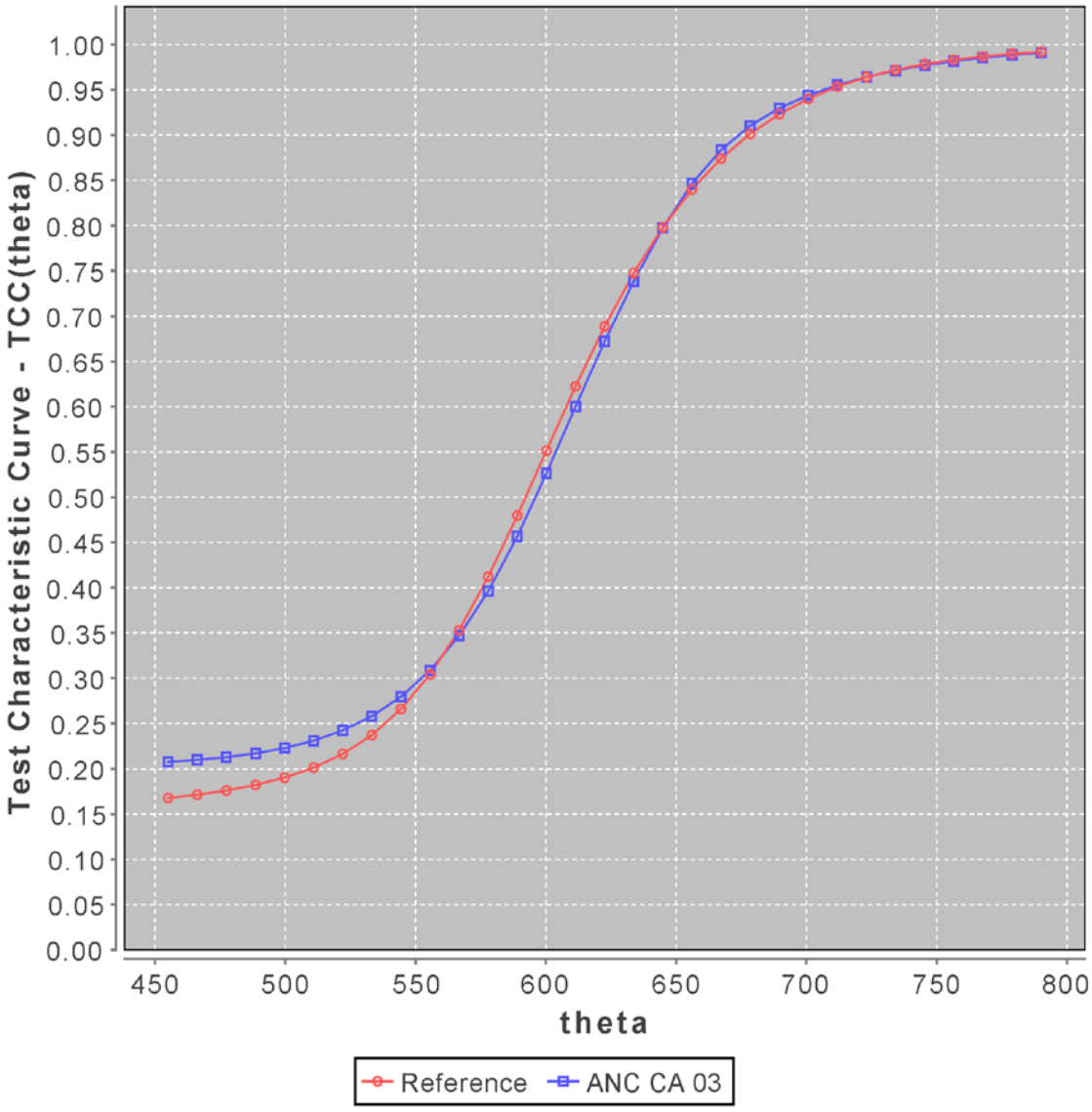

Figure 15 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 3 Communication Arts

MO Anchor Grade 4 CA TCC Plot from Form Construction

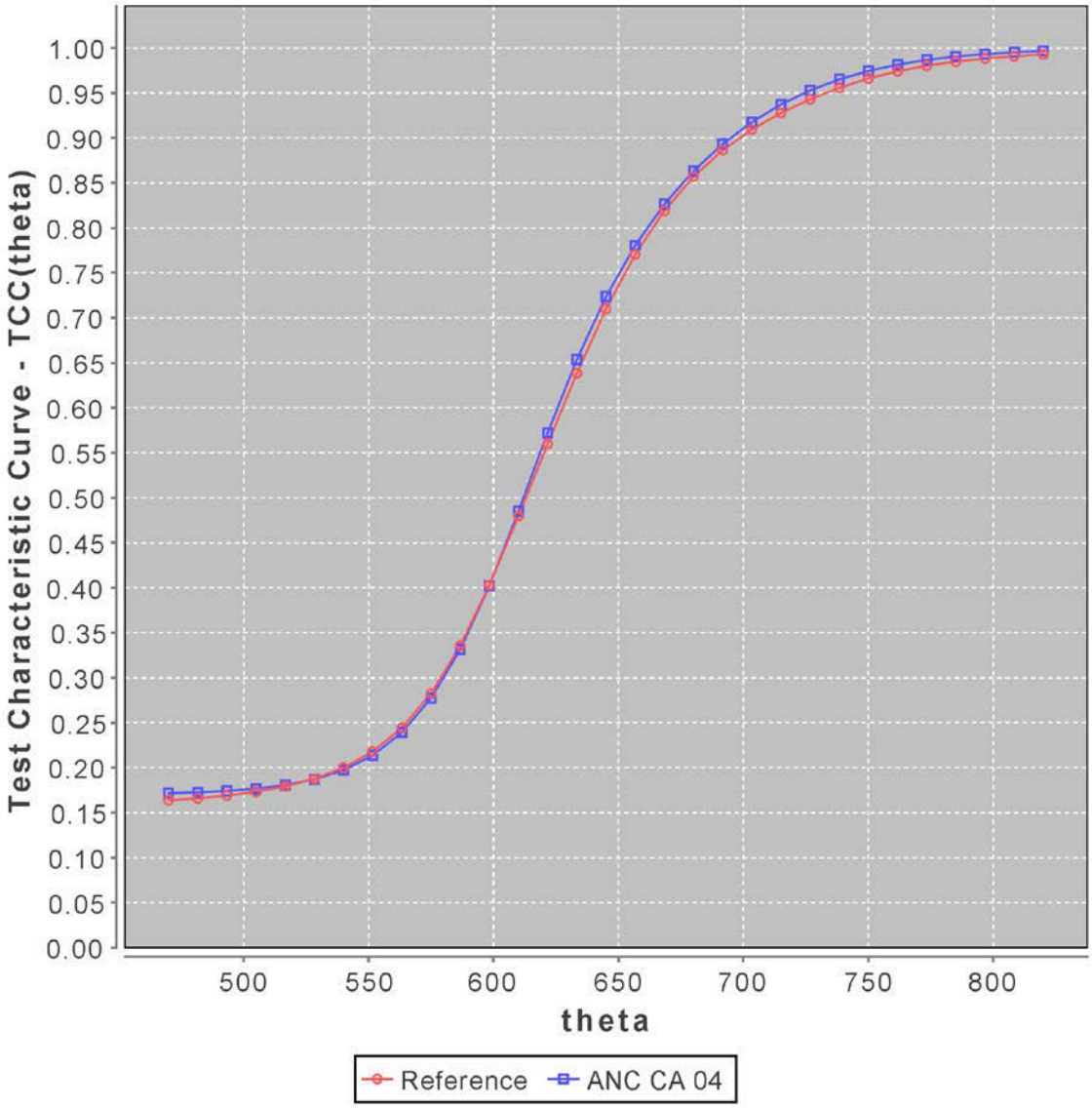

Figure 16 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 4 Communication Arts

MO Anchor Grade 5 CA TCC Plot from Form Construction

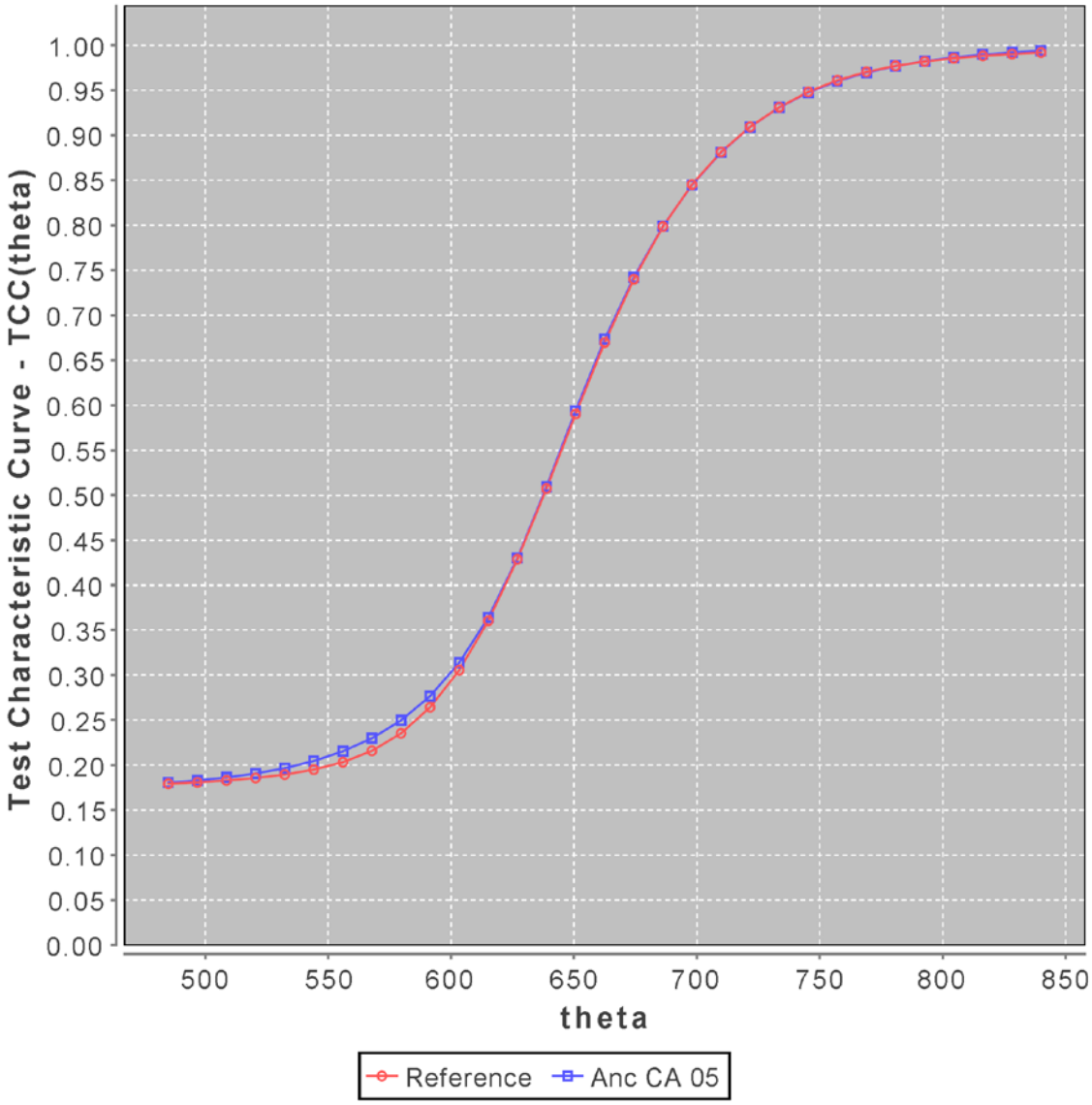

Figure 17 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 5 Communication Arts

MO Anchor Grade 6 CA TCC Plot from Form Construction

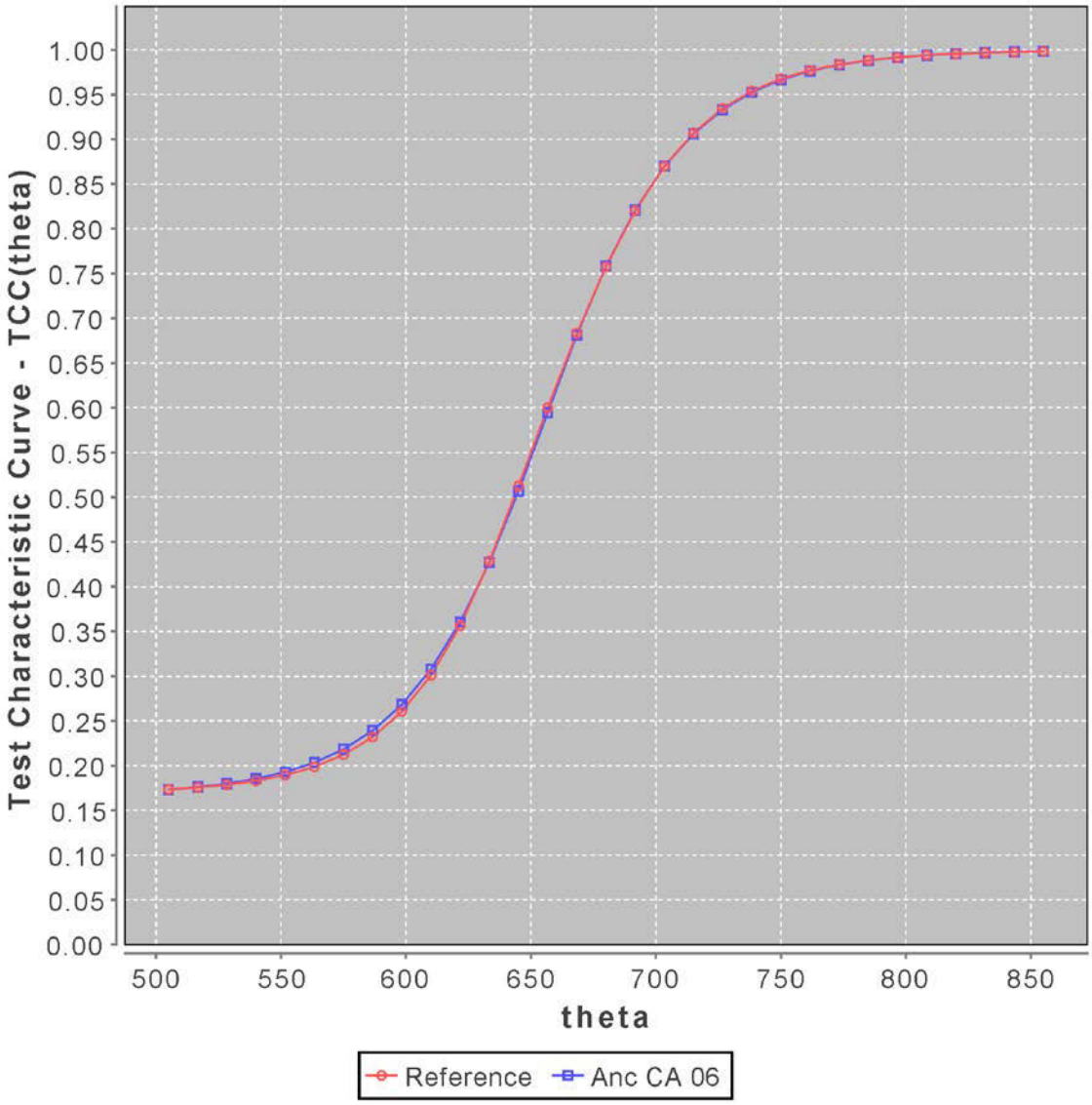

Figure 18 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 6 Communication Arts

MO Anchor Grade 7 CA TCC Plot from Form Construction

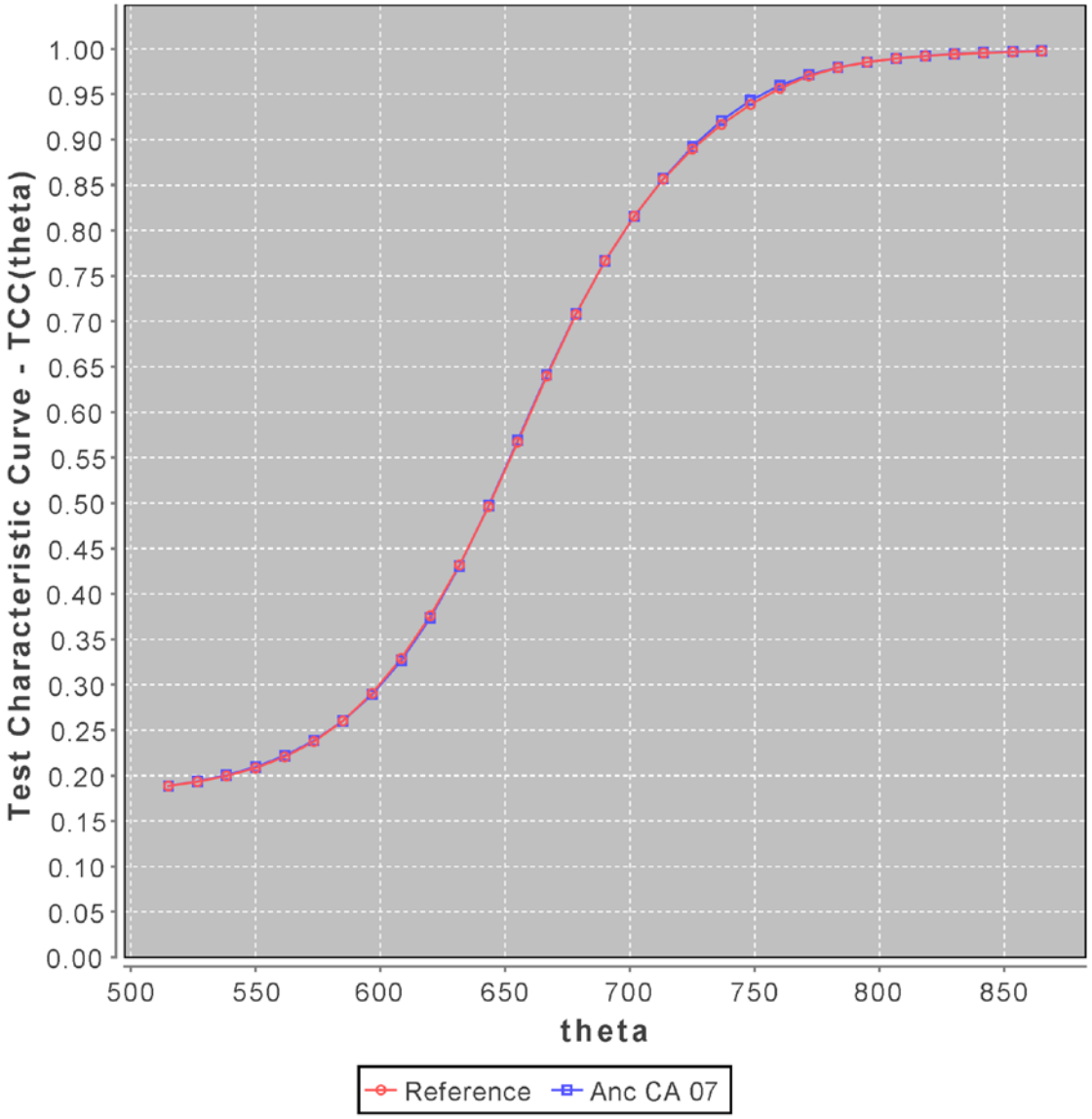

Figure 19 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 7 Communication Arts

MO Anchor Grade 8 CA TCC Plot from Form Construction

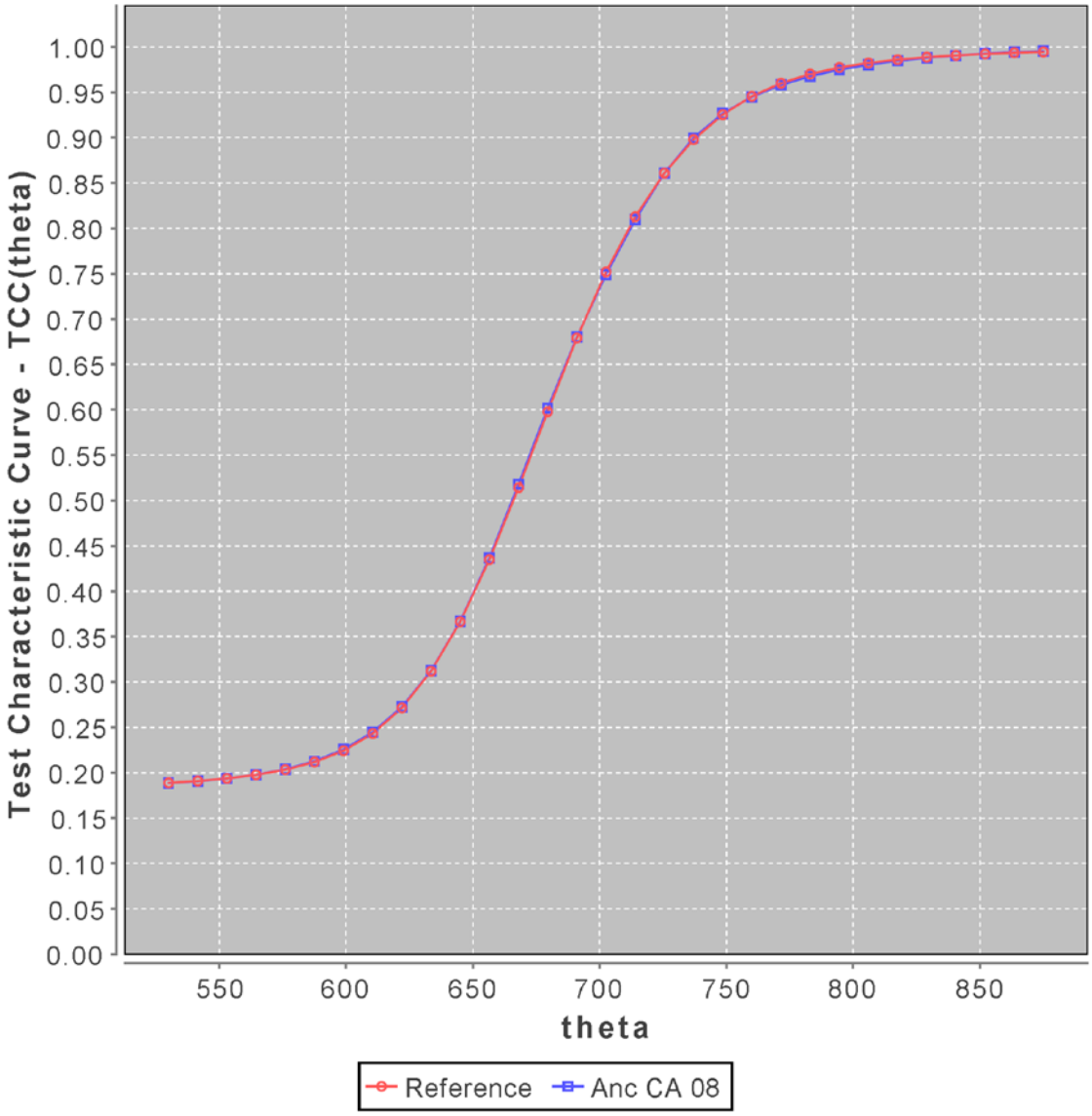

Figure 20 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 8 Communication Arts

MO Anchor Grade 3 MA TCC Plot from Form Construction

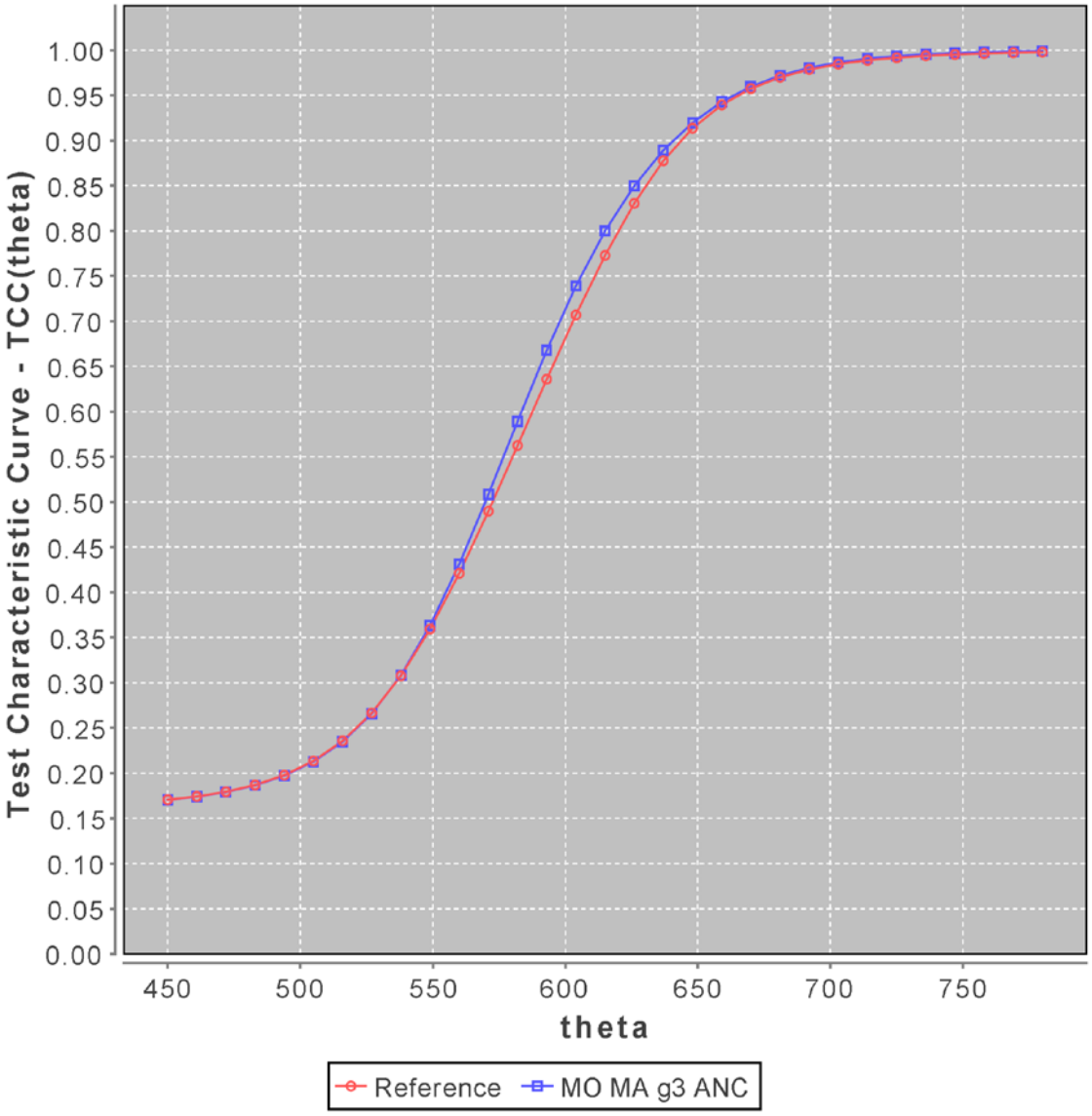

Figure 21 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 3 Mathematics

MO Anchor Grade 4 MA TCC Plot from Form Construction

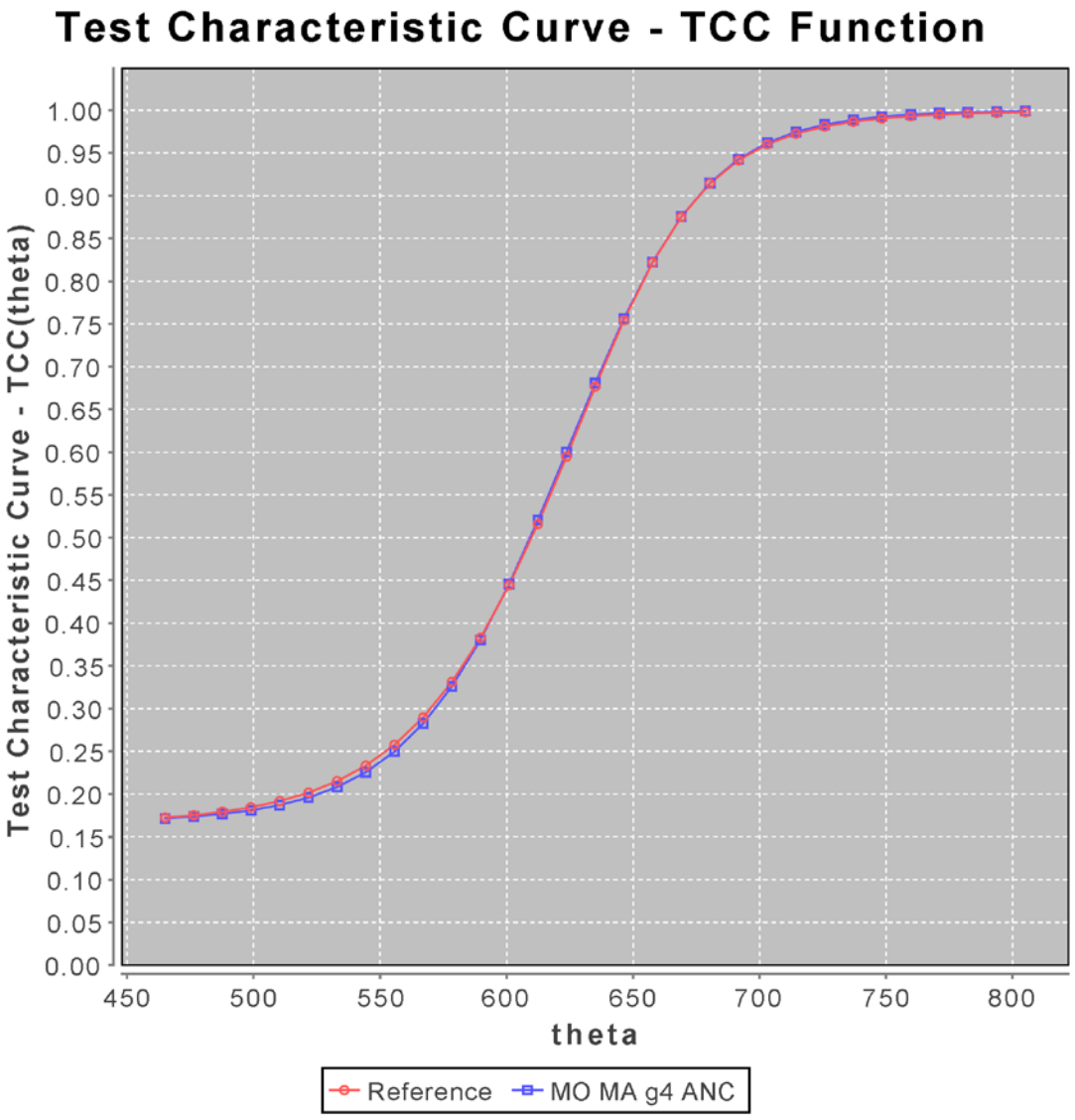

Figure 22 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 4 Mathematics

MO Anchor Grade 5 MA TCC Plot from Form Construction
*Not provided

MO Anchor Grade 6 MA TCC Plot

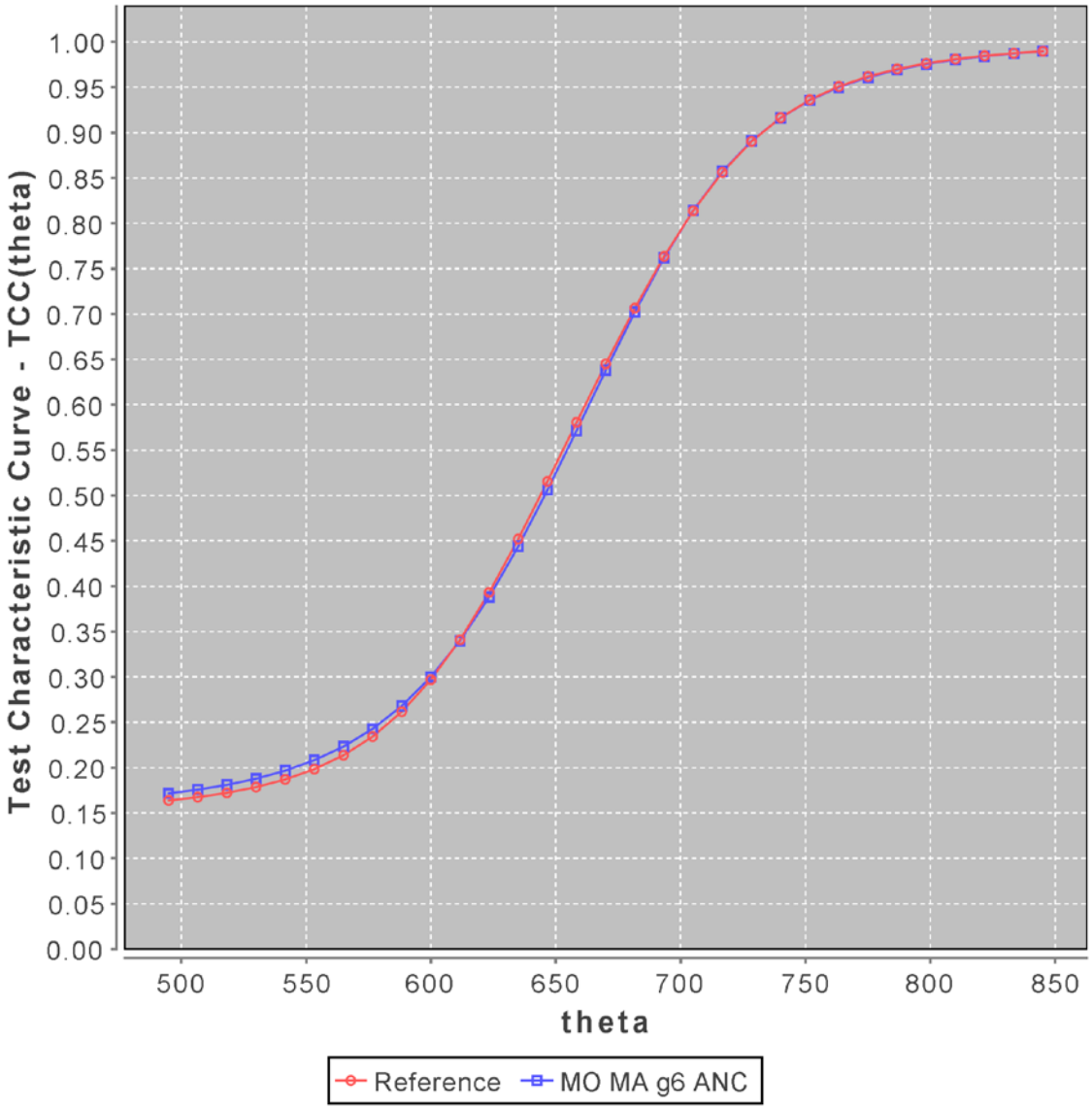

Figure 23 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 6 Mathematics

MO Anchor Grade 7 MA TCC Plot from Form Construction

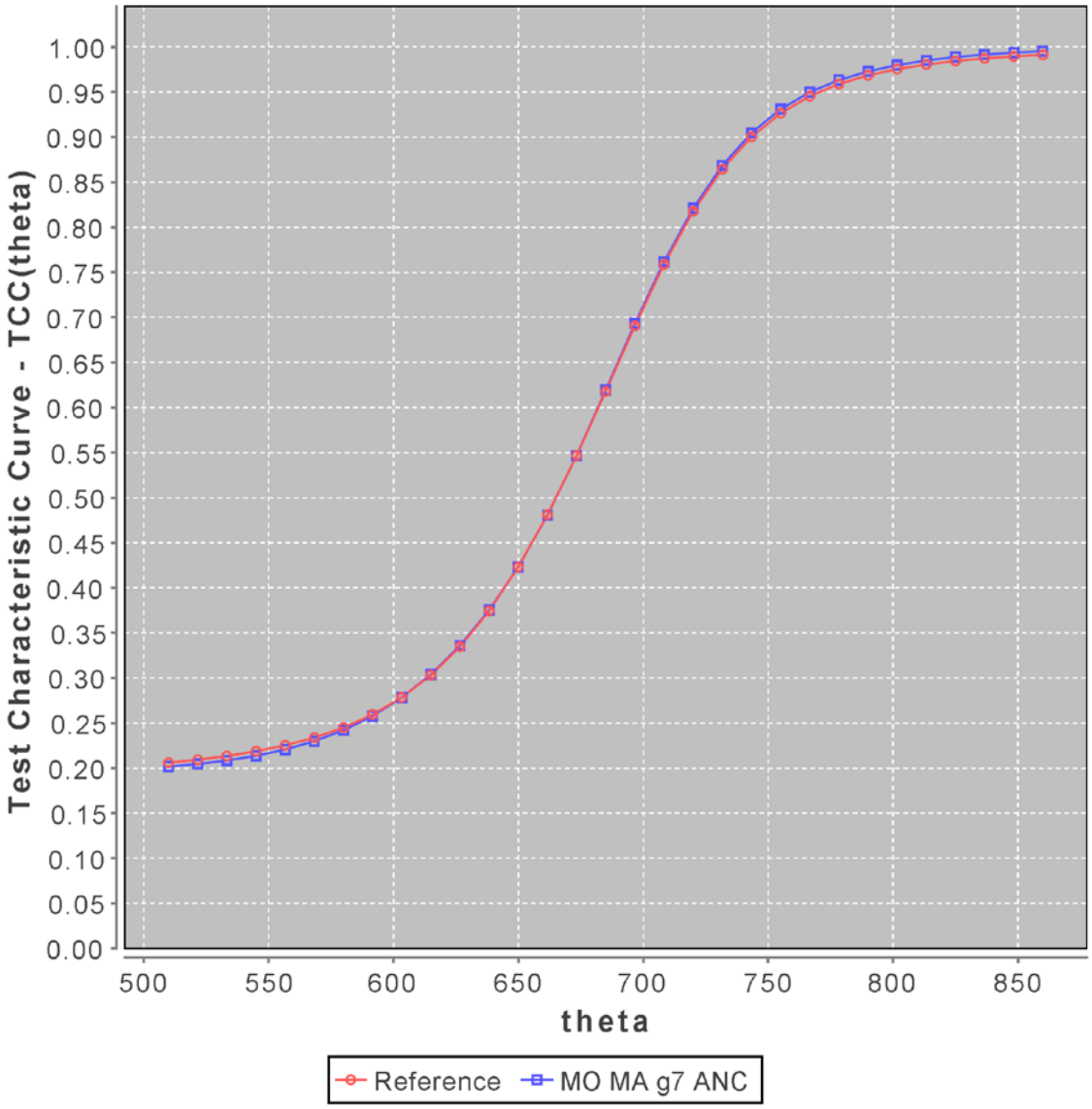

Figure 24 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 7 Mathematics

MO Anchor Grade 8 MA TCC Plot from Form Construction

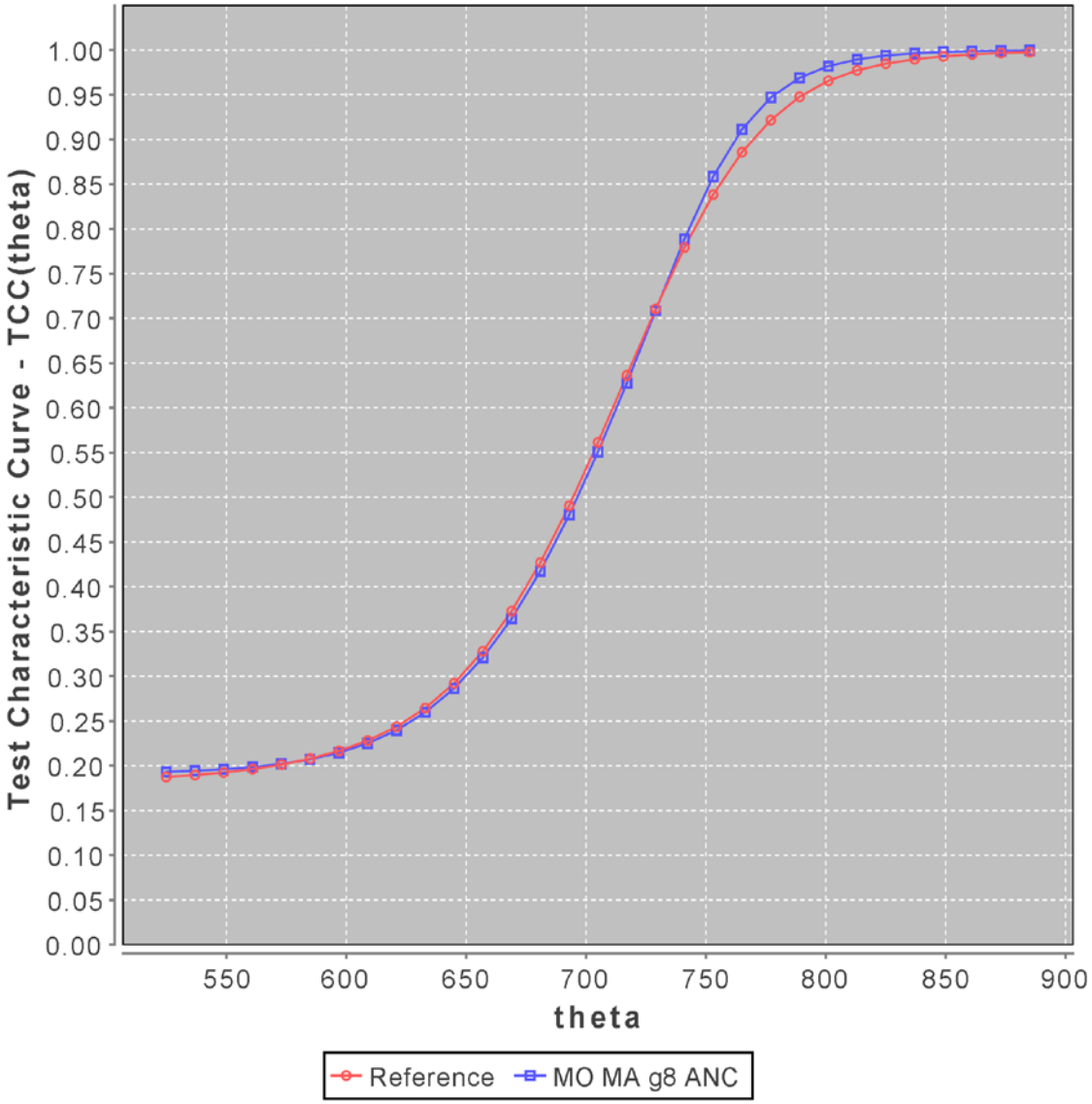

Figure 25 MAP Anchor Test Characteristic Curves for Selected Anchor Compared to 2012 Reference Anchor, Grade 8 Mathematics

MO Anchor Grade 5 SC TCC Plot from Form Construction
*Not Provided
MO Anchor Grade 8 SC TCC Plot from Form Construction
*Not Provided

Anchor Construct

The anchor set was selected to match the 2012 anchor blueprint. In other words, the distribution of items measuring each content standard should be similar in 2012 and 2014. CTB provided a rule that the distribution of items may vary by 10 percentage points in either direction. DESE provided CTB's blueprint tables used

during form construction. These tables showed an "anchor target" which was the percentage of items that was supposed to measure each content standard, and the percentage of items measuring each content standard in the selected anchor.

Table 25 and Table 26 show the difference in the percentage of items measuring each content standard between the selected anchor and the anchor target for Communication Arts and Mathematics. All percentages are within the specified range, except for Grade 4 Communication Arts. The percentages are outside of the range specified by CTB. In the anchor target, 70% (or 8 items) of the items are supposed measure Reading. In the selected anchor, 58% (or 7 items) of the items measured Reading. Given the one item difference, it seems unlikely that this shift in construct caused the drop in performance on the MAP in Grade 4 Communication Arts.

Since the Grade 5 Mathematics anchor and both Science anchors changed from the original selected anchors, these tables are not available.

Table 25 Average Change in Percentage of Anchor Items Measuring Each Content Standard in 2014 compared to 2013, Communication Arts

	3	4	5	6	7	8
Speaking/Writing Standard English	3%	12%	3%	8%	6%	8%
Reading—Fiction & Nonfiction	-4%	-12%	-4%	-8%	-6%	-8%
Writing Formally & Informally	0%	0%	0%	0%	0%	0%

Table 26 Average Change in Percentage of Anchor Items Measuring Each Content Standard in 2014 compared to 2013, Mathematics

	3	4	5	6	7	8
Number and Operations	-2%	0%		-7%	-1%	-2%
Algebraic Relationships	1%	1%		3%	-2%	3%
Geometric and Spatial Relationships	-3%	-1%		0%	3%	-2%
Measurement	6%	1%		0%	1%	-2%
Data and Probability	-2%	-3%		5%	-3%	3%

Anchor Placement

Anchor items should be placed in the same relative location each time that they are used. Items changing location on the test can adversely affect student performance on anchor items.

Using the provided item metadata files, the original location of each anchor was compared to its location in 2014. Table 27 shows the frequency distribution of the number of locations anchor items moved from their previous placement on a MAP form. For Communication Arts and Mathematics, all items moved between +/-5 spots. For Grade 5 Science, 18 of the items did not move from their previous location, and four of the items moved more than 20 places in the test book. For Grade 8 Science, 20 of the 25 anchor items changed location by more than 20 places. The remaining 5 items change location between 10 and 19 places from their previous administration.

Test booklets were not provided; therefore, we could not study the presentation of the items in the test books.

Table 27 Frequency Distribution of Number of Locations Anchor Items Moved from Previous Administration for all Grades and Content Areas

Con-tent	Grade	# of Anc	Number of Locations Item Moved from Previous Location												
			-5	-4	-3	-2	-1	0	1	2	3	4	5	10-19	20+
CA	G3	12						12							
	G4	11*					1	9	1						
	G5	12					1	10		1					
	G6	12*						12							
	G7	14			2		1	10	1						
	G8	13					2	10		1					
MA	G3	13						9		4					
	G4	14						13				1			
	G5	11				1		5			1	1	3		
	G6	13*	1		2		5	1	1	1	1	1			
	G7	14				1	3	5	3	1	1				
	G8	13		1		1	1	4		3	1	2			
SC	G5	22						18							4
	G8	25												5	20

*Previous administration of one item could not be determined

Stocking and Lord Equating

CTB uses the Stocking and Lord (1983) procedure to link the current year's assessment to the existing scale. DESE provided the results from the Stocking and Lord procedure used to put the 2014 assessment on the MAP test scale.

Figure 15 through Figure 25 show the input TCCs and estimate TCCs. The input TCCs are based on the anchor items' parameters from the previous administration. The estimate TCCs use transformed estimates from the current calibration. Figure 15 through Figure 25 show that the input and TCCs are very similar to one another.

One means of evaluating the success of the Stocking and Lord procedure is to examine the correlation between the input and estimated parameters. Missouri has established procedures (see page B-2 of the 2013 Missouri Technical Report) where the correlation of input and estimate *b*-parameters should be at or above .90 and the correlation of input and estimate *a*-parameters should be at or above .80.

Table 28 shows the correlation of the input and estimate *a*- and *b*-parameters for all grades and content areas. All correlations are well above the previously established thresholds.

Table 28 Correlation of Anchor Items Input and Estimate A- and B-Parameters for All Grades and Content Areas

Content Area	Grade	Number of Items	Correlation A-Parameter	Correlation B-Parameter
Communication Arts	3	12	0.99	0.97
	4	12	0.84	1.00
	5	12	0.99	0.98
	6	13	0.98	0.95
	7	14	0.96	0.97
	8	13	0.96	0.99
Mathematics	3	13	0.97	0.98
	4	14	0.96	0.92
	5	11	0.99	0.97
	6	14	0.99	0.96
	7	14	0.95	0.98
	8	13	0.97	0.99
Science	5	22	0.96	0.98
	8	25	0.94	0.97

Empirical Evaluation of the Anchor Items

Finally, the anchor items themselves were examined to compare their performance with previous assessments. In particular, we examined p-values and differences between input and estimated TCCs

Anchor P-Values

We examined p-values in two ways. First, we looked at the 2014 and previous p-values for all anchor items. Then, for the grades with large drops in scores (Grades 3 and 4 Communication Arts and Grade 4 Mathematics), we identified all common items from 2013 to see if performance on those items also declined in 2014.

Table 29 and Table 30 show the mean p-values for the 2014 anchor compared to the average of their previous p-values for Communication Art and Mathematics, respectively. The data from the previous p-values for Science were not available for this analysis since the anchor set changed between the time the anchor was selected and the anchor was used. For Grade 5 Mathematics, this comparison is only based on the 10 items from the original anchor.

Table 29 and Table 30 show that, in general, the mean p-values were similar to the average of their previous values. There was one item on the Grade 4 Mathematics test which had a notable change in p-value from the previous form. The p-value for this item was 0.51 in 2014, and it was 0.69 the previous time it was administered. Without this item, the average difference would have been 0.0 between 2014 and the previous administration.

Table 29 Mean P-values of Anchor Items in 2014 and in Previous Administration, Communication Arts

Grade	Number of Anchors	2014	SD	Previous	SD	Average Difference (2014-Previous)	SD
3	12	0.74	0.14	0.74	0.14	0.00	0.03
4	12	0.76	0.19	0.76	0.17	-0.01	0.02
5	12	0.72	0.10	0.71	0.10	0.01	0.02
6	13	0.69	0.12	0.69	0.12	0.00	0.03
7	14	0.71	0.16	0.70	0.17	0.02	0.04
8	13	0.68	0.15	0.68	0.14	0.00	0.02

Table 30 Mean P-values of Anchor Items in 2014 and in Previous Administration, Mathematics

Grade	Number of Anchors	2014	SD	Previous	SD	Average Difference (2014-Previous)	SD
3	13	0.81	0.09	0.80	0.10	0.01	0.02
4	14	0.72	0.14	0.73	0.12	-0.01	0.05
5	10	0.75	0.10	0.74	0.12	0.01	0.03
6	14	0.70	0.14	0.69	0.13	0.00	0.04
7	14	0.65	0.15	0.64	0.13	0.01	0.04
8	13	0.55	0.18	0.59	0.18	-0.04	0.02

For those tests with a large decline in performance, we compared performance on items (regardless of anchor status) from the 2013 administration. The purpose of this analysis was to see how students performed on those items that were administered in both 2013 and 2014. It was hoped that an examination of common items between 2013 and 2014 would provide evidence to support or disconfirm the decline in test performance. Unfortunately there were very few items that were common to both 2013 and 2014.

Table 31 shows the results of this analysis for the limited number of items that were available. In Grade 3 Communication Arts, performance on the four common items did not change between 2013 and 2014. In Grade 4 Communication Arts, performance on the two common items did decline from 2013. In Grade 4 Mathematics, performance on the three common items declined from 2013. There were so few items available for this analysis that it is inappropriate to draw firm conclusions about supporting or disconfirming evidence from this analysis.

Table 31 Mean P-values of Items Common Between 2013 and 2014

Grade	Content Area	Number of Items	2014 average	2013 average
3	Communication Arts	4	0.68	0.68
4	Communication Arts	2	0.63	0.67
4	Mathematics	3	0.66	0.67

Anchor Item Characteristic Curves (ICCs)

DESE provided CTB's output for the anchor evaluation conducted each year. Appendix B in the 2013 Missouri Technical Report details the analyses approved by the Missouri Technical Advisory Committee. The following paragraph from the Missouri Technical Report summarizes the TAC-approved steps for evaluating anchor items.

Items removed from the anchor set are still scored as part of the whole test. Anchor

items are considered for exclusion from the MAP[2] under the following conditions[3]:

1. *Items flagged using the TCC method are considered for exclusion when the correlation between the input and equated item parameters is below .90 for the b-parameter or below .80 for the a-parameter. If the exclusion of an outlying anchor item increases the correlation to above .90 for the b-parameter or above .80 for the a-parameter, then the anchor is a candidate for removal.*
2. *An item is a candidate for removal when it is flagged for large differences on four of the seven statistics considered when examining the differences between the IRT regression curves.*
3. *Removal of the item will only be considered after alternative explanations have been considered that may explain shifts in performance. For example, performance on the anchor item may improve because of a statewide initiative emphasizing instruction on a particular set of skills. In this case, improved performance on the item represents true growth in that area. Removing the anchor item may artificially lower test scores*
4. *Removal of the item may not significantly alter the content distribution of the anchor set. The distribution of the anchor items across the content standards must remain within 10% of the MAP test blueprint.*
5. *The number of remaining items will remain at an acceptable level of anchor set reliability. Operationally, this means the anchor set will still be representative of the total test blueprint and that the anchor may not be less than 20% of the total test length.*

Using these criteria, we examined the anchor items. As mentioned in the previous section, the input and estimated parameters were highly correlated. In this section, we examine the difference between input and estimated curves (#2 in the evaluation criteria). The seven statistics include (again from the 2013 Missouri Technical Report):

- *UnWtd Mean = Average signed difference in estimated probability.*
- *UnWtd Mean Abs Dif = Average Absolute (unsigned) difference in estimated probability.*
- *UnWtd RMSD = Root mean squared difference.*
- *Wtd Mean = Weighted average signed difference in estimated probability.*
- *Wtd Mean Abs = Weighted average Absolute (unsigned) difference in estimated*
- *probability.*
- *WtdRMSD = Weighted Root mean squared difference.*

For the six statistics listed above, differences greater than +.10 are considered large, and differences between +.07 and .10 are considered moderate.

---

[2] The anchor would be excluded from the anchor set. Students would still be scored on their performance on the item.

[3] DESE should specify if these criteria are sequential or conditional.

Additionally, the Maximum Absolute difference (MaxAbsDifPC) will be identified. For MaxAbsDIFPC, large differences are those greater than +.15, and moderate differences are all differences between +.125 and .15.

Using these criteria, all anchor items were examined. Table 32, Table 33, and Table 34 show item fit statistics for all grades within Communication Arts, Mathematics, and Science. All, except three, met the evaluation criteria. One Grade 4 Mathematics item, one Grade 7 Communication Arts item, and one Grade 7 Mathematics item was flagged for possible removal from the internal anchor. Table 32 shows the values on each statistic. Figure 26 shows the input and estimated item characteristic curves (ICCs) for the Grade 4 Mathematics item, Figure 27 shows the input and estimated ICCs for Grade 7 Communication Arts item, and Figure 28 shows the input and estimated ICCs for Grade 7 Mathematics item.

Grade 4 Mathematics. This item was identified for possible removal during the July 28 to August 1 window. During this window, DESE examined the anchor item for possible differences in presentation between the time it was first presented in the 2011 assessment and the 2014 assessment. Similar analyses should be completed for the Grade 7 Communication Arts and Mathematics items. The item had the same placement on the page, and it was presented in the same relative item position. DESE reported no differences in the item format.

Next, DESE examined the item to see if it covered a skill that was not part of the Grade 4 Mathematics Common Core State Standards. The item's content aligns to the Grade 4 Mathematics Common Core State Standards, so Missouri students should be taught the concepts measured by the item using both Missouri Grade-level Expectations and the Mathematics Common Core State Standards.

Table 32 Values on Anchor Evaluation Statistics for Communication Arts

Gd	ItemId	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
3	1	0.02	0.02	0.03	-0.02	0.03	0.03	-0.03
	2	0.02	0.02	0.04	-0.01	0.02	0.01	-0.01
	3	0.01	0.00	0.01	0.00	0.00	0.00	0.00
	4	0.01	0.01	0.03	-0.01	0.02	0.01	-0.01
	5	0.01	0.00	0.01	0.00	0.01	0.01	0.01
	6	0.06	0.05	0.09	0.05	0.08	0.08	0.08
	7	0.01	0.01	0.02	0.00	0.01	0.01	-0.01
	8	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	9	0.01	0.00	0.01	0.00	0.00	0.00	0.00
	10	0.02	0.01	0.04	-0.01	0.02	0.02	-0.02
	11	0.00	0.00	0.01	0.00	0.01	0.00	0.00
	12	0.00	0.00	0.01	0.00	0.00	0.00	0.00
4	1	0.02	0.02	0.04	-0.02	0.03	0.03	-0.03
	2	0.02	0.02	0.04	0.00	0.02	0.02	-0.01
	3	0.03	0.02	0.05	0.00	0.03	0.02	0.01
	4	0.01	0.01	0.02	0.00	0.01	0.01	0.01
	5	0.02	0.01	0.04	0.01	0.02	0.02	0.02
	6	0.01	0.01	0.03	0.01	0.01	0.01	0.00
	7	0.03	0.02	0.07	0.02	0.02	0.01	0.01
	8	0.02	0.02	0.04	-0.01	0.02	0.01	-0.01
	9	0.01	0.01	0.03	-0.01	0.01	0.01	0.00
	10	0.01	0.01	0.02	0.00	0.01	0.01	0.00
	11	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	12	0.01	0.00	0.01	0.00	0.00	0.00	0.00

Gd	ItemId	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
5	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.02	0.02	0.04	-0.01	0.02	0.01	0.00
	3	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	4	0.01	0.00	0.01	0.00	0.01	0.01	0.01
	5	0.02	0.02	0.04	0.02	0.03	0.02	0.02
	6	0.01	0.01	0.02	0.00	0.01	0.01	0.00
	7	0.02	0.02	0.03	-0.02	0.02	0.02	-0.02
	8	0.02	0.01	0.02	-0.01	0.01	0.01	-0.01
	9	0.02	0.01	0.03	-0.01	0.03	0.02	-0.02
	10	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
	11	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	12	0.01	0.01	0.02	0.01	0.02	0.01	0.01
6	1	0.05	0.04	0.08	-0.04	0.06	0.06	-0.06
	2	0.03	0.02	0.06	0.01	0.02	0.01	0.00
	3	0.01	0.00	0.01	0.00	0.01	0.01	0.01
	4	0.03	0.02	0.06	-0.02	0.05	0.04	-0.04
	5	0.01	0.01	0.02	-0.01	0.02	0.02	-0.02
	6	0.02	0.01	0.03	0.01	0.02	0.01	0.01
	7	0.02	0.01	0.03	-0.01	0.02	0.02	-0.02
	8	0.01	0.01	0.02	0.01	0.02	0.02	0.02
	9	0.00	0.00	0.01	0.00	0.01	0.01	0.01
	10	0.05	0.05	0.08	0.05	0.07	0.06	0.06
	11	0.01	0.01	0.03	0.00	0.02	0.02	0.02
	12	0.01	0.00	0.01	0.00	0.01	0.01	0.01
	13	0.02	0.01	0.02	0.01	0.02	0.02	0.02

Gd	ItemId	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
7	1	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	2	0.06	0.04	0.11	-0.04	0.05	0.03	-0.03
	3	0.02	0.02	0.03	-0.01	0.02	0.01	-0.01
	4	0.04	0.03	0.07	0.03	0.05	0.05	0.05
	5	0.10	0.06	0.21	0.06	0.15	0.13	0.13
	6	0.02	0.02	0.04	-0.02	0.03	0.03	-0.03
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.02	0.01	0.04	-0.01	0.03	0.02	-0.02
	9	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
	10	0.03	0.02	0.04	-0.02	0.03	0.03	-0.03
	11	0.02	0.02	0.03	-0.02	0.03	0.03	-0.03
	12	0.01	0.01	0.02	0.00	0.01	0.01	0.00
	13	0.01	0.00	0.02	0.00	0.01	0.01	-0.01
	14	0.01	0.01	0.02	-0.01	0.02	0.02	-0.02
8	1	0.01	0.01	0.02	0.00	0.02	0.02	0.01
	2	0.05	0.04	0.08	-0.03	0.05	0.05	-0.05
	3	0.02	0.02	0.03	0.00	0.02	0.02	0.00
	4	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	5	0.03	0.02	0.07	0.02	0.04	0.04	0.04
	6	0.01	0.01	0.02	0.00	0.02	0.01	0.01
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.01	0.01	0.02	0.00	0.01	0.01	0.00
	9	0.02	0.02	0.04	0.01	0.02	0.01	0.01
	10	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
	11	0.02	0.02	0.04	0.01	0.02	0.01	0.00
	12	0.01	0.01	0.01	0.00	0.01	0.00	0.00
	13	0.02	0.01	0.06	0.00	0.03	0.02	0.00

Table 33 Values on Anchor Evaluation Statistics for Mathematics

Gd	ItemId	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
3	1	0.02	0.01	0.04	0.01	0.02	0.01	0.01
	2	0.00	0.00	0.01	0.00	0.01	0.00	0.00
	3	0.02	0.02	0.06	-0.02	0.03	0.02	-0.02
	4	0.02	0.01	0.03	0.01	0.02	0.02	0.02
	5	0.01	0.01	0.02	0.01	0.01	0.01	0.01
	6	0.01	0.01	0.02	0.00	0.01	0.01	0.00
	7	0.01	0.00	0.01	0.00	0.00	0.00	0.00
	8	0.03	0.02	0.05	-0.02	0.04	0.04	-0.04
	9	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	10	0.02	0.02	0.04	0.01	0.01	0.01	0.00
	11	0.01	0.01	0.02	0.01	0.01	0.01	0.01
	12	0.03	0.02	0.06	-0.02	0.03	0.03	-0.03
	13	0.01	0.01	0.03	0.00	0.02	0.01	0.01
4	1	0.03	0.02	0.05	-0.02	0.03	0.02	-0.02
	2	0.01	0.01	0.02	0.00	0.01	0.01	0.01
	3	0.03	0.02	0.05	0.02	0.04	0.04	0.04
	4	0.01	0.01	0.03	0.00	0.02	0.02	-0.01
	5	0.02	0.01	0.06	0.01	0.04	0.03	0.03
	6	0.02	0.02	0.03	0.02	0.03	0.03	0.03
	7	0.06	0.04	0.12	0.04	0.05	0.04	0.04
	8	0.01	0.01	0.03	-0.01	0.02	0.02	-0.02
	9	0.01	0.01	0.01	0.00	0.01	0.00	0.00
	10	0.02	0.01	0.03	0.00	0.01	0.01	0.01
	11	0.01	0.01	0.02	0.01	0.01	0.01	0.01
	12	0.01	0.01	0.03	0.00	0.02	0.02	0.02
	13	0.12	0.10	0.19	-0.10	0.16	0.15	-0.15
	14	0.01	0.01	0.02	0.00	0.01	0.01	0.00

Gd	ItemId	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
5	1	0.01	0.01	0.03	0.01	0.02	0.02	0.02
	2	0.03	0.03	0.05	0.00	0.02	0.02	0.00
	3	0.01	0.01	0.03	0.01	0.02	0.02	0.02
	4	0.04	0.03	0.07	0.03	0.06	0.06	0.06
	5	0.01	0.01	0.01	0.00	0.01	0.00	0.00
	6	0.02	0.01	0.05	-0.01	0.02	0.01	-0.01
	7	0.04	0.04	0.07	-0.04	0.05	0.05	-0.05
	8	0.01	0.01	0.03	-0.01	0.02	0.01	-0.01
	9	0.02	0.02	0.04	-0.02	0.03	0.03	-0.03
	10	0.01	0.00	0.02	0.00	0.01	0.01	0.01
	11	0.01	0.01	0.02	0.00	0.01	0.01	0.00
6	1	0.01	0.01	0.03	-0.01	0.02	0.02	-0.02
	2	0.05	0.04	0.09	-0.04	0.07	0.07	-0.07
	3	0.02	0.02	0.05	0.02	0.02	0.02	0.02
	4	0.02	0.02	0.03	-0.02	0.03	0.03	-0.03
	5	0.02	0.02	0.03	0.01	0.02	0.02	0.01
	6	0.01	0.00	0.01	0.00	0.01	0.01	0.01
	7	0.02	0.02	0.04	-0.02	0.02	0.02	-0.02
	8	0.01	0.01	0.02	0.01	0.01	0.01	0.01
	9	0.01	0.01	0.03	-0.01	0.02	0.02	-0.02
	10	0.01	0.01	0.03	-0.01	0.01	0.01	-0.01
	11	0.06	0.04	0.10	0.04	0.05	0.04	0.04
	12	0.01	0.01	0.02	0.00	0.01	0.01	0.00
	13	0.01	0.01	0.02	0.00	0.01	0.01	0.00
	14	0.05	0.04	0.10	0.04	0.08	0.08	0.08

Gd	ItemId	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
7	1	0.01	0.01	0.02	0.01	0.02	0.02	0.02
	2	0.01	0.01	0.03	-0.01	0.02	0.02	-0.02
	3	0.03	0.02	0.05	-0.01	0.02	0.01	0.01
	4	0.02	0.01	0.04	0.00	0.02	0.02	-0.01
	5	0.08	0.05	0.21	0.05	0.12	0.10	0.10
	6	0.02	0.01	0.03	-0.01	0.02	0.02	-0.01
	7	0.04	0.04	0.07	-0.04	0.06	0.06	-0.06
	8	0.03	0.02	0.05	0.02	0.02	0.02	0.02
	9	0.02	0.01	0.03	-0.01	0.02	0.02	-0.02
	10	0.01	0.01	0.03	0.00	0.02	0.01	-0.01
	11	0.02	0.02	0.04	-0.01	0.03	0.03	-0.03
	12	0.02	0.02	0.04	-0.02	0.03	0.03	-0.03
	13	0.03	0.02	0.05	0.02	0.04	0.04	0.04
	14	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
8	1	0.01	0.01	0.03	0.01	0.02	0.02	0.02
	2	0.01	0.01	0.02	0.01	0.01	0.01	0.01
	3	0.02	0.01	0.04	-0.01	0.02	0.02	-0.02
	4	0.04	0.02	0.09	0.02	0.06	0.05	0.05
	5	0.01	0.01	0.02	0.00	0.01	0.01	-0.01
	6	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01
	7	0.01	0.01	0.02	0.00	0.01	0.01	0.00
	8	0.04	0.02	0.08	-0.02	0.05	0.04	-0.04
	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	10	0.03	0.02	0.05	-0.02	0.03	0.02	-0.02
	11	0.01	0.01	0.02	0.00	0.01	0.01	0.01
	12	0.01	0.01	0.02	0.00	0.01	0.01	0.01
	13	0.02	0.01	0.04	0.01	0.02	0.01	0.01

Table 34 Values on Anchor Evaluation Statistics for Science

Gd	ItemId	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
5	1	0.06	0.04	0.10	-0.04	0.04	0.03	-0.03
	2	0.01	0.00	0.01	0.00	0.01	0.00	0.00
	3	0.01	0.01	0.03	-0.01	0.01	0.00	0.00
	4	0.01	0.01	0.02	0.01	0.01	0.01	0.01
	5	0.01	0.00	0.01	0.00	0.01	0.01	-0.01
	6	0.04	0.03	0.07	0.03	0.03	0.02	0.02
	7	0.06	0.04	0.12	0.04	0.05	0.03	0.03
	8	0.01	0.01	0.03	0.00	0.00	0.00	0.00
	9	0.03	0.02	0.06	-0.02	0.03	0.02	-0.02
	10	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	11	0.06	0.06	0.07	-0.06	0.06	0.06	-0.06
	12	0.01	0.01	0.01	0.00	0.01	0.01	0.00
	13	0.01	0.01	0.02	-0.01	0.01	0.01	0.00
	14	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
	15	0.00	0.00	0.01	0.00	0.01	0.01	-0.01
	16	0.03	0.02	0.06	0.02	0.04	0.04	0.04
	17	0.02	0.01	0.03	-0.01	0.03	0.03	-0.03
	18	0.02	0.01	0.05	0.01	0.03	0.03	0.03
	19	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	20	0.03	0.02	0.07	-0.02	0.05	0.04	-0.04
	21	0.04	0.03	0.10	0.02	0.07	0.06	0.06
	22	0.00	0.00	0.01	0.00	0.00	0.00	0.00

Gd	ItemId	UnWtd RMSD	UnWtd Mean Abs Difference	UnWtd Max	UnWtd Mean	Wtd RMSD	Wtd Mean Abs Difference	Wtd Mean
8	1	0.01	0.01	0.03	0.00	0.01	0.01	0.00
	2	0.02	0.01	0.04	-0.01	0.01	0.01	0.00
	3	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.03	0.02	0.07	0.02	0.04	0.03	0.03
	6	0.01	0.01	0.01	0.00	0.01	0.01	-0.01
	7	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
	8	0.04	0.04	0.06	-0.04	0.05	0.04	-0.04
	9	0.01	0.01	0.02	-0.01	0.02	0.02	-0.02
	10	0.02	0.02	0.03	-0.02	0.03	0.03	-0.03
	11	0.02	0.01	0.04	0.01	0.02	0.02	0.02
	12	0.03	0.02	0.08	0.02	0.05	0.05	0.05
	13	0.03	0.03	0.06	0.03	0.05	0.05	0.05
	14	0.00	0.00	0.01	0.00	0.01	0.01	0.01
	15	0.05	0.03	0.12	0.03	0.08	0.06	0.06
	16	0.02	0.01	0.05	0.01	0.04	0.03	0.03
	17	0.05	0.03	0.10	-0.03	0.07	0.07	-0.07
	18	0.01	0.01	0.02	-0.01	0.01	0.01	-0.01
	19	0.00	0.00	0.01	0.00	0.01	0.01	0.00
	20	0.02	0.02	0.03	0.02	0.03	0.03	0.03
	21	0.03	0.03	0.07	-0.02	0.05	0.05	-0.04
	22	0.05	0.04	0.09	0.04	0.06	0.06	0.06
	23	0.03	0.03	0.06	0.01	0.02	0.01	0.01
	24	0.06	0.04	0.13	-0.04	0.09	0.08	-0.08
	25	0.02	0.02	0.04	-0.02	0.03	0.02	-0.02

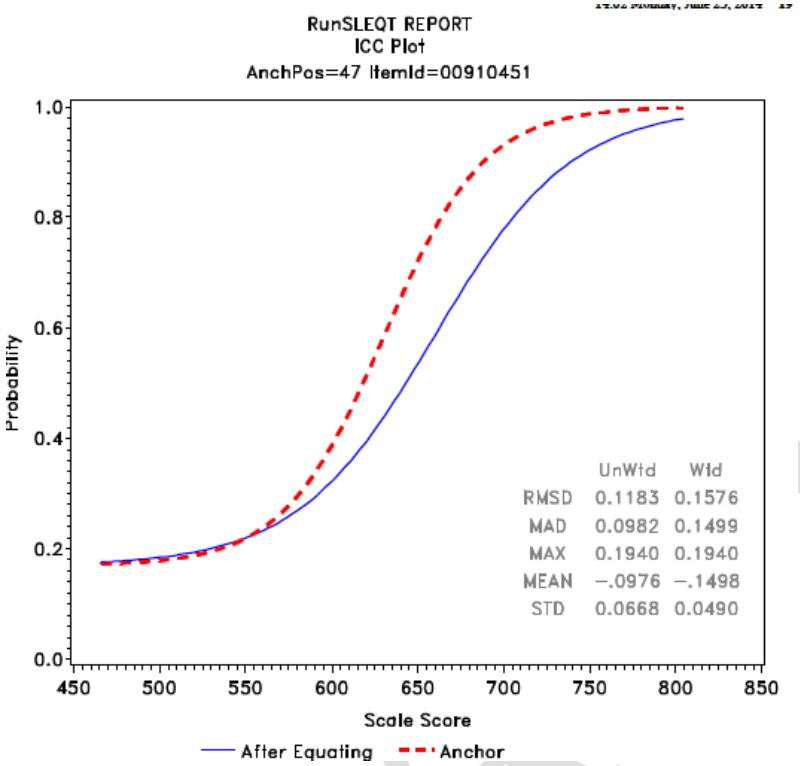

Figure 26. Flagged Item from Grade 4 Mathematics that was not removed from internal anchor

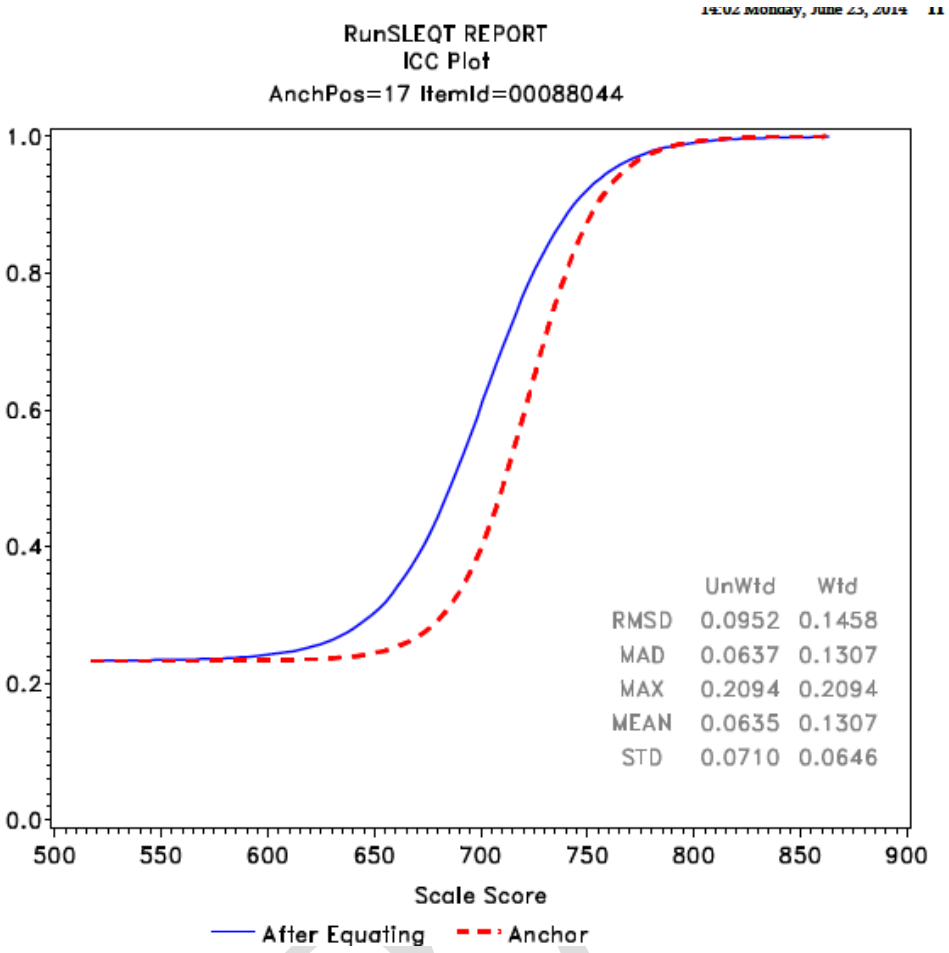

Figure 27. Flagged Item from Grade 7 Communication Arts that was not removed from internal anchor

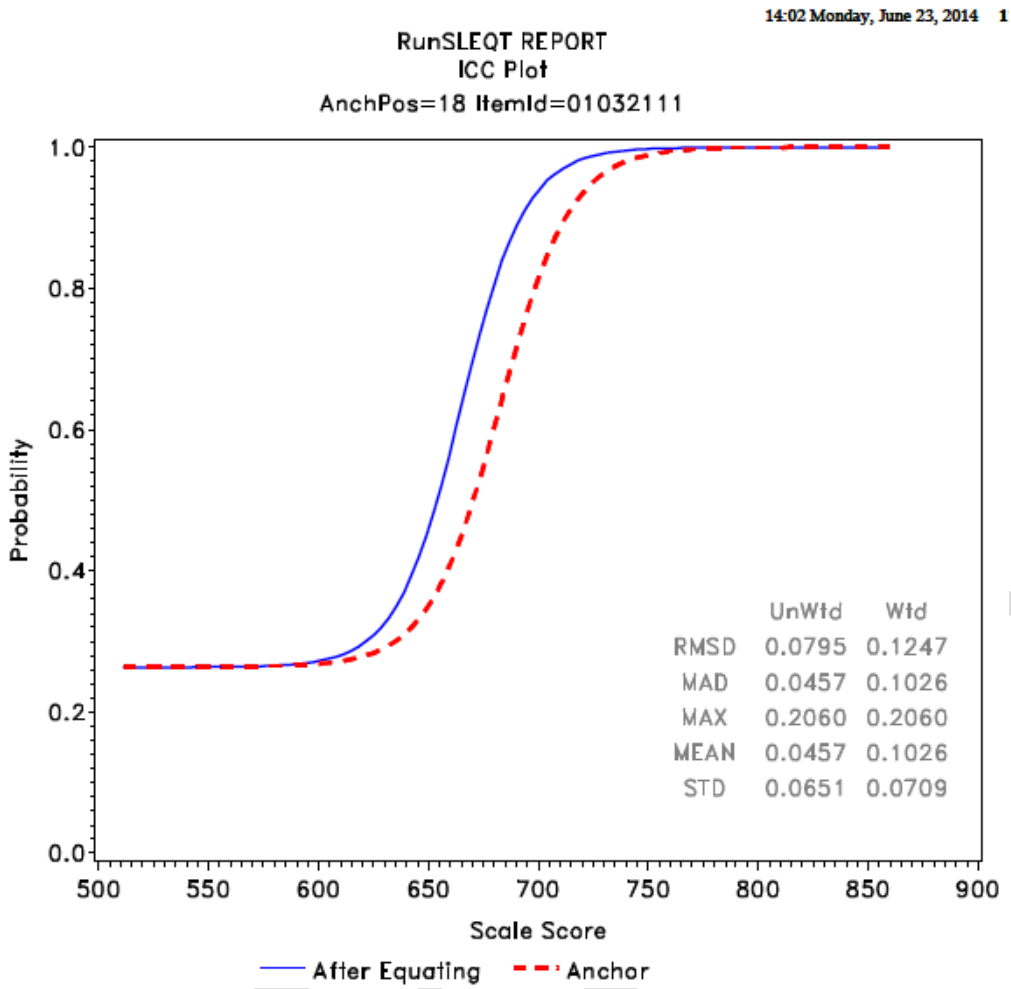

Figure 28. Flagged Item from Grade 7 Mathematics that was not removed from internal anchor

Summary of the Anchor Comparability

The 2014 internal anchor appears to be comparable in terms of test construct and psychometric characteristics to the internal anchor used for the previous MAP form. In most grades and content areas the majority of the 2014 anchor was selected from the 2009 form. The 2014 measures a similar construct with similar psychometric characteristics compared to the previous form. The results form the Stocking and Lord equating indicate that the equating was sound. Closer examination of the anchor items themselves revealed that three items that could be considered for removal from the anchor set. While the removal of each of these items may have affected the outcome of the equating in each of their particular grades, idiosyncratic performance on anchor items does not appear to be the primary cause of the decline in performance.

Summary and Conclusions

This work was precipitated by the decline in performance on the 2014 MAP compared to previous years. This paper examined the comparability of the:

- Test population
- Test construct
- Test administration
- Psychometric characteristics
- Internal anchor

No analysis pointed to a single reason for the decline in test performance. There does not appear to be one consistent reason for the decline in scores; rather, there appears to be a host of reasons working together that contribute to the overall decline in scores. First, the administration conditions of the test this year varied from past years. Many schools missed a significant number of snow days. A handful of schools participated in Smarter Balanced field testing. There was a good deal of anti-testing sentiment in the state. Second, this is a transition year for many schools in terms of teaching the CCSS versus the GLEs. Other states also saw a decline in test scores this year as their school transitioned from state content standards to CCSS. Third, Missouri had an almost complete replacement of its test form for the first time since 2010.

It is routine for a testing program to change or refresh test forms on an ongoing basis. Whenever forms are changed or refreshed, it increases the challenge of comparing test scores from year to year. This is why new test forms are carefully constructed to align to the same blueprint with similar difficulty as previous test forms. In Missouri, financial constraints had prevented the state to making changes to the test form since 2010. Prior to 2010, the test form was completely renewed every other year. There were sound logistical and technical reasons for making changes to the test and the internal anchor in 2014; however, those changes add to the complexity of attempting to explain the change in performance between 2013 and 2014. Even though these analyses cannot consistently explain the decline in MAP performance, neither did they reveal any systematic issues or irregularities in the administration or processing or results that would explain the decline in performance, lending support to the integrity of the results.