MI Access
Michigan's Alternate Assessment Program

Functional Independence

Accessing Print, Expressing Ideas, Mathematics, and Science

Michigan's Alternate Assessment Program

Michigan Department of Education
Bureau of Assessment and Accountability

Technical Report
2011-2012 Addendum

October 2012

Compiled by
Adam E. Wyse

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

The MI–Access Technical Reports provide information about (a) the nature of the tests; (b) their intended uses; (c) the processes involved in their development; (d) technical information related to scoring, interpretation, and evidence of reliability and validity; (e) scaling and equating; and (f) guidelines for test administration and interpretation, as recommended by the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999, p. 67). Technical Reports have been developed for the Functional Independence assessments and the Participation/Supported Independence assessments.

The following Technical Reports have been developed:

Functional Independence ELA/Mathematics, March 2007
Participation and Supported Independence ELA/Mathematics, June 2007
Participations/Supported Independence/Functional Independence Science, August 2008

Each year, an addendum will be produced to provide the technical quality evidence for the most recent operational administrations of the tests. This is the sixth annual addendum and includes the Functional Independence (FI) Accessing Print, Expressing Ideas, Mathematics, and Science tests administered in the 2011 – 2012 school year.

As indicated in the full technical reports for MI–Access, the reports are designed to communicate with multiple users, including state policy makers and their staffs, school and district administrators, teachers, and parents and other advocates interested in such documentation. However, the addendums are designed to provide annual technical quality updates for a much smaller audience. The addendums will focus on reliability and validity evidence gathered at the time of test administration, scoring and equating, and reporting.

1. Form Design

The form design of the 2011 – 2012 operational tests was unchanged from last year's design except for FI Science grade 5 where there was one fewer Reflecting item and one more Earth Science item this year compared to last year. Tables 1.1 to 1.4 contain the test blueprints. In ELA (i.e., Accessing Print and Expressing Ideas), four forms were developed. In Mathematics and Science, two forms were developed.

Each form also contained a set of anchor items that were used to facilitate equating to the score scale originally developed in 2005 – 2006 for ELA and Mathematics, and in 2007 – 2008 for Science. Anchor items were included among the core items as they counted toward the total score. The number of equating items is also included in Tables 1.1 to 1.4

Table 1.1
Operational Mathematics Test Blueprint Grades 3 to 8

Strand	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Numbers and Operations	10	16	16	18	17	17
Algebra						2
Measurement	8	8	10	12	12	10
Geometry	9	4	2	2	3	3
Data and Probability	3	2	2	3	3	3
Total Core Items	30	30	30	35	35	35
Equating Items	8	8	8	8	8	8
Embedded Field-test Items	8	8	8	10	10	10
Total Test Items	38	38	38	45	45	45

Table 1.2
Operational Mathematics Test Blueprint Grade 11

Strand	Grade 11
Patterns and Relationships	4
Geometry and Measurement	16
Data analysis and Statistics	2
Number Sense and Numeration	15
Numerical and Algebraic Operations	3
Total Core Items	40
Equating Items	10
Embedded Field-test Items	10
Total Test Items	50

Table 1.3
Operational English Language Arts Test Blueprint Grades 3 to 11

Strand	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
Word Recognition	20	20	20	20	20	20	20
Text Comprehension							
Narrative Text	7	7	7	7	7	7	7
Informational Text	7	7	7	7	7	7	7
Functional Text	7	7	7	7	7	7	7
Accessing Print Total	41	41	41	41	41	41	41
Expressing Ideas Prompt	0	1	0	0	1	0	1
Total Core Items	41	42	41	41	42	41	42
Equating Items	11	11	11	11	11	11	11
Embedded Field-test Items	11	12	11	11	12	11	12
Total Test Items	52	54	52	52	54	52	54

Table 1.4

Operational Science Test Blueprint Grades 5, 8, and 11

Strand	Grade 5	Grade 8	Grade 11
Constructing & Reflecting	2	2	2
Reflecting	1	2	2
Life Science	13	14	14
Physical Science	12	14	15
Earth Science	7	8	12
Total Core Items	35	40	45
Equating Items	8	10	10
Embedded Field-test Items	8	10	10
Total Test Items	43	50	55

2. Participation

Participation in the assessments is monitored by gender, by racial/ethnic group, by economically disadvantaged, and by accommodation usage. These student-level characteristics are also used to evaluate differential item functioning (DIF) when the groups are large enough to support the analysis. These results are reported in Section 8.

Participation counts and percentages by gender and grade are given in Tables 2.1 – 2.4 for Accessing Print, Expressing Ideas, Mathematics, and Science, respectively, participation counts and percentages by race/ethnicity and grade are given in Tables 2.5 – 2.8, participation counts and percentages by economically disadvantaged and grade are given in Tables 2.9 – 2.12, and participation counts and percentages by accommodation usage and grade are given in Tables 2.13 – 2.16. In general, there were roughly twice as many males as females. The largest racial/ethnic group was White students with 59.2% to 66.0% of the students, followed by Black students with 24.8% to 31.0% of the students, Hispanic students with 4.0% to 8.0% of the students, and Asian with about 1% of the students. In general, there were approximately two to three times as many economically disadvantaged students as non-economically disadvantaged students. The largest accommodation was for Reader (e.g., Read aloud accommodation) with 30.4% to 69.2% of the students. The highest percentage was at grade 3 and it became progressively smaller as grade level increased. The next largest percentage was for Audio CD with 6.0% to 14.0% of the students, followed by Other accommodations with 0.8% to 8.2% of the students. The use of the Audio CD accommodation tended to be higher in middle school and high school compared to elementary school and the use of Other accommodations tended to be less in middle school and high school. For Expressing Ideas at grades 4, 7, and 11, where students were administered a prompt, 14.3%, 8.2%, and 4.3% of the students, respectively, had a scribe as an accommodation. The Braille and Large Print accommodations were used by a very small proportion of students in each grade.

Table 2.1
2011–2012 N-Counts and Percents by Gender and Grade for Accessing Print

Grade	Female		Male		Total
	N	%	N	%	N
3	495	32.8	1013	67.2	1508
4	587	33.6	1159	66.4	1746
5	664	35.2	1221	64.8	1885
6	661	35.3	1209	64.7	1870
7	659	35.3	1207	64.7	1866
8	671	36.5	1167	63.5	1838
11	633	38.4	1112	61.6	1650

Table 2.2

2011–2012 N-Counts and Percents by Gender and Grade for Expressing Ideas

Grade	Female		Male		Total
	N	%	N	%	N
4	582	33.7	1143	66.3	1725
7	652	35.4	1189	64.5	1841
11	630	38.7	998	61.3	1628

Table 2.3

2011–2012 N-Counts and Percents by Gender and Grade for Mathematics

Grade	Female		Male		Total
	N	%	N	%	N
3	469	34.7	884	65.3	1353
4	554	35.7	997	64.3	1551
5	636	37.0	1085	63.0	1721
6	658	36.3	1153	63.7	1811
7	672	36.8	1156	63.2	1828
8	688	37.6	1143	62.4	1831
11	639	38.4	1023	61.6	1662

Table 2.4

2011–2012 N-Counts and Percents by Gender and Grade for Science

Grade	Female		Male		Total
	N	%	N	%	N
5	629	36.0	1119	64.0	1748
8	689	36.8	1185	63.2	1874
11	640	38.5	1021	61.5	1661

Table 2.5

2011–2012 N-Counts and Percents by Ethnicity and Grade for Accessing Print

Grade	Native Hawaiian or Other Pacific Islander		American Indian or Alaskan Native		Black or African American		Hispanic of any race		White		Multi-racial		Asian		Total
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N
3	0	0.0	22	1.5	374	24.8	120	8.0	929	61.6	40	2.7	23	1.5	1508
4	1	0.1	20	1.1	442	25.3	123	7.0	1100	63.0	47	2.7	13	0.7	1746
5	3	0.2	23	1.2	510	27.1	132	7.0	1159	61.5	41	2.2	17	0.9	1885
6	0	0.0	30	1.6	516	27.6	124	6.6	1140	61.0	36	1.9	24	1.3	1870
7	1	0.1	25	1.3	557	29.8	115	6.2	1128	60.5	23	1.2	17	0.9	1866
8	2	0.1	15	0.8	550	29.9	83	4.5	1137	61.9	29	1.6	22	1.2	1838
11	2	0.1	24	1.5	425	25.8	67	4.1	1086	65.8	27	1.6	19	1.2	1650

Table 2.6

2011–2012 N-Counts and Percents by Ethnicity and Grade for Expressing Ideas

Grade	Native Hawaiian or Other Pacific Islander		American Indian or Alaskan Native		Black or African American		Hispanic of any race		White		Multi-racial		Asian		Total
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N
4	1	0.1	19	1.1	428	24.8	121	7.0	1096	63.5	47	2.7	13	0.8	1725
7	1	0.1	25	1.4	545	29.6	113	6.1	1117	60.7	23	1.2	17	0.9	1841
11	2	0.1	23	1.4	416	25.6	67	4.1	1074	66.0	28	1.7	18	1.1	1628

Table 2.7

2011–2012 N-Counts and Percents by Ethnicity and Grade for Mathematics

Grade	Native Hawaiian or Other Pacific Islander		American Indian or Alaskan Native		Black or African American		Hispanic of any race		White		Multi-racial		Asian		Total
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N
3	0	0.0	18	1.3	357	26.4	98	7.2	818	60.5	40	3.0	22	1.6	1353
4	0	0.0	15	1.0	418	27.0	118	7.6	951	61.3	39	2.5	10	0.6	1551
5	3	0.2	16	0.9	477	27.7	117	6.8	1052	61.1	38	2.2	18	1.0	1721
6	0	0.0	27	1.5	519	28.7	114	6.3	1095	60.5	36	2.0	20	1.1	1811
7	1	0.1	24	1.3	566	31.0	110	6.0	1083	59.2	26	1.4	18	1.0	1828
8	2	0.1	15	0.8	560	30.6	87	4.8	1117	61.0	29	1.6	21	1.1	1831
11	2	0.1	24	1.4	431	25.9	68	4.1	1092	65.7	26	1.6	19	1.1	1794

Table 2.8

2011–2012 N-Counts and Percents by Ethnicity and Grade for Science

Grade	Native Hawaiian or Other Pacific Islander		American Indian or Alaskan Native		Black or African American		Hispanic of any race		White		Multi-racial		Asian		Total
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N
5	3	0.2	18	1.0	498	28.5	117	6.7	1056	60.4	37	2.1	19	1.1	1748
8	2	0.1	14	0.7	572	30.5	89	4.7	1144	61.0	32	1.7	21	1.1	1874
11	2	0.1	24	1.4	429	25.8	67	4.0	1094	65.9	26	1.6	19	1.1	1661

Table 2.9

2011-2012 N-Counts and Percents by Economically Disadvantaged and Grade for Accessing Print

Grade	Economically Disadvantaged		Non-Economically Disadvantaged		Total
	N	%	N	%	N
3	1125	74.6	383	25.4	1508
4	1305	74.7	441	25.3	1746
5	1372	72.8	513	27.2	1885
6	1355	72.5	515	27.5	1870
7	1355	72.6	511	27.4	1866
8	1324	72.0	514	28.0	1838
11	1107	67.1	543	32.9	1788

Table 2.10

2011-2012 N-Counts and Percents by Economically Disadvantaged and Grade for Expressing Ideas

Grade	Economically Disadvantaged		Non-Economically Disadvantaged		Total
	N	%	N	%	N
4	1289	74.7	436	25.3	1725
7	1339	72.7	502	27.3	1841
11	1091	67.0	537	33.0	1628

Table 2.11

2011-2012 N-Counts and Percents by Economically Disadvantaged and Grade for Mathematics

Grade	Economically Disadvantaged		Non-Economically Disadvantaged		Total
	N	%	N	%	N
3	1014	74.9	339	25.1	1353
4	1179	76.0	372	24.0	1551
5	1241	72.1	480	27.9	1721
6	1299	71.7	512	28.3	1811
7	1336	73.1	492	26.9	1828
8	1315	71.8	516	28.2	1831
11	1117	67.2	545	32.8	1662

Table 2.12

2011-2012 N-Counts and Percents by Economically Disadvantaged and Grade for Science

Grade	Economically Disadvantaged		Non-Economically Disadvantaged		Total
	N	%	N	%	N
5	1266	72.4	482	27.6	1748
8	1343	71.7	531	28.3	1874
11	1116	67.2	545	32.8	1661

Table 2.13
2011–2012 N-Counts and Percents by Accommodation and Grade for Accessing Print

Grade	Reader		Audio		Braille		Enlarged Print		Other		Scribe		Word Processed		All Students
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N
3	959	63.6	90	6.0	0	0.0	3	0.2	88	5.8	NA		NA		1508
4	1044	59.8	129	7.4	1	0.1	12	0.7	120	6.9	NA		NA		1746
5	1017	54.0	203	10.8	1	0.1	3	0.2	155	8.2	NA		NA		1885
6	916	49.0	195	10.4	1	0.1	6	0.3	105	5.6	NA		NA		1870
7	818	43.8	215	11.5	1	0.1	5	0.3	66	3.5	NA		NA		1866
8	822	44.7	226	12.3	4	0.2	6	0.3	162	3.4	NA		NA		1838
11	558	33.8	180	10.9	8	0.5	7	0.4	20	1.2	NA		NA		1650

Table 2.14
2011–2012 N-Counts and Percents by Accommodation and Grade for Expressing Ideas

Grade	Reader		Audio		Braille		Enlarged Print		Other		Scribe		Word Processed		All Students
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N
4	1035	60.0	128	7.4	1	0.1	11	0.6	84	4.9	246	14.3	11	0.6	1725
7	813	44.2	212	11.5	1	0.1	6	0.3	66	3.6	147	8.0	25	1.4	1841
11	555	34.1	177	10.9	8	0.5	7	0.4	13	0.8	69	4.2	14	0.9	1628

Table 2.15
2011–2012 N-Counts and Percents by Accommodation and Grade for Mathematics

Grade	Reader		Audio		Braille		Enlarged Print		Other		Scribe		Word Processed		All Students
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N
3	936	69.2	111	8.2	0	0.0	5	0.4	77	5.7	NA		NA		1353
4	1012	65.2	147	9.5	1	0.1	12	0.8	88	5.7	NA		NA		1551
5	1032	60.0	241	14.0	2	0.1	2	0.1	137	8.0	NA		NA		1721
6	943	52.1	215	11.9	2	0.1	5	0.3	105	5.8	NA		NA		1811
7	835	45.7	218	11.9	1	0.1	6	0.3	71	3.9	NA		NA		1828
8	836	45.7	238	13.0	6	0.3	6	0.3	55	3.0	NA		NA		1831
11	506	30.4	180	10.8	8	0.5	6	0.4	21	1.3	NA		NA		1662

Table 2.16

2011–2012 N-Counts and Percents by Accommodation and Grade for Science

Grade	Reader		Audio		Braille		Enlarged Print		Other		Scribe		Word Processed		All Students
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N
5	1084	62.0	230	13.2	3	0.2	3	0.2	98	5.6	NA		NA		1748
8	879	46.9	263	14.0	6	0.3	5	0.3	50	2.7	NA		NA		1874
11	536	32.3	178	10.7	7	0.4	7	0.4	18	1.1	NA		NA		1661

Form Distribution

Recall from Section 1, four forms were developed for Accessing Print at grades 3 – 8 and 11 and Expressing Ideas grades 4, 7, and 11, and two forms were developed for Mathematics at grades 3 – 8 and 11 and for Science at each of the three grades. These forms were distributed to districts and schools according to the guidelines from the Michigan Department of Education Bureau of Assessment and Accountability. The sampling unit was the school. Forms were randomly assigned using stratified random sampling where stratification was based on the enrollment counts provided to Questar. Except for Detroit, each district received up to two forms at a grade. For Detroit, each school received the same form. One additional condition was imposed on the distribution of forms. Due to cost considerations, only Form 1 was developed for the Audio CD, Braille, and Enlarged Print accommodations. Hence, all students with these accommodations were administered Form 1. Tables 2.17, 2.18, and 2.19 show the materials ordered versus the number of forms that were used to test the students based on the sampling plan. The number of materials ordered usually exceeded the number tested by a somewhat considerable margin. In general, the number of students tested was larger on Form 1 than the other forms due to the fact students that got the Audio CD, Braille, or Enlarged Print had to receive Form 1.

The percent of students by various subgroups and form for the 2011 – 2012 school year are given in Tables 2.20, 2.21, 2.22, and 2.23. Each table contains the number of students tested by form at each grade, as well as the grade total. At each grade, the percent of students for the various subgroups is given by form as well as for the grade total. The percents for Accessing Print, Expressing Ideas, Mathematics, and Science are given in Tables 2.20, 2.21, 2.22 and 2.23, respectively. The subgroups consist of gender, three racial/ethnic groups (Black, Hispanic, and White), and three other subgroups (Economically Disadvantaged, English Language Learners or ELL, and Formerly Limited English Proficient or FLEP). As seen from the tables for all three content areas, each form was well represented by the various subgroups. Moreover, for each form at a grade, the percent of students across the subgroups was generally consistent with the percents for the grade population.

Table 2.17

2011–2012 Material Orders versus Test Forms Used for Accessing Print & Expressing Ideas

Assessment	Ordered	Tested
FI Accessing Print Grade 3 Form 1	1362	525
FI Accessing Print Grade 3 Form 2	881	300
FI Accessing Print Grade 3 Form 3	975	337
FI Accessing Print Grade 3 Form 4	1082	374
FI Accessing Print & Expressing Ideas Grade 4 Form 1	1625	610
FI Accessing Print & Expressing Ideas Grade 4 Form 2	1102	352
FI Accessing Print & Expressing Ideas Grade 4 Form 3	1134	426
FI Accessing Print & Expressing Ideas Grade 4 Form 4	1203	414
FI Accessing Print Grade 5 Form 1	1722	702
FI Accessing Print Grade 5 Form 2	1118	436
FI Accessing Print Grade 5 Form 3	1147	398
FI Accessing Print Grade 5 Form 4	1124	378
FI Accessing Print Grade 6 Form 1	1552	653
FI Accessing Print Grade 6 Form 2	990	441
FI Accessing Print Grade 6 Form 3	1008	404
FI Accessing Print Grade 6 Form 4	990	403
FI Accessing Print & Expressing Ideas Grade 7 Form 1	1606	724
FI Accessing Print & Expressing Ideas Grade 7 Form 2	1001	392
FI Accessing Print & Expressing Ideas Grade 7 Form 3	955	383
FI Accessing Print & Expressing Ideas Grade 7 Form 4	990	409
FI Accessing Print Grade 8 Form 1	1528	673
FI Accessing Print Grade 8 Form 2	976	440
FI Accessing Print Grade 8 Form 3	882	380
FI Accessing Print Grade 8 Form 4	886	362
FI Accessing Print Grade 11 Form 1	1320	701
FI Accessing Print Grade 11 Form 2	812	379
FI Accessing Print Grade 11 Form 3	814	306
FI Accessing Print Grade 11 Form 4	824	328

Table 2.18
2011–2012 Material Orders versus Test Forms Used for Mathematics

Assessment	Ordered	Tested
FI Mathematics Grade 3 Form 1	2315	828
FI Mathematics Grade 3 Form 2	1719	555
FI Mathematics Grade 4 Form 1	2571	873
FI Mathematics Grade 4 Form 2	2079	713
FI Mathematics Grade 5 Form 1	2720	1030
FI Mathematics Grade 5 Form 2	2099	717
FI Mathematics Grade 6 Form 1	2445	1051
FI Mathematics Grade 6 Form 2	1896	786
FI Mathematics Grade 7 Form 1	2487	1046
FI Mathematics Grade 7 Form 2	1945	808
FI Mathematics Grade 8 Form 1	2364	1059
FI Mathematics Grade 8 Form 2	1827	794
FI Mathematics Grade 11 Form 1	2092	1034
FI Mathematics Grade 11 Form 2	1591	677

Table 2.19
2011–2012 Material Orders versus Test Forms Used for Science

Assessment	Ordered	Tested
FI Science Grade 5 Form 1	2752	1011
FI Science Grade 5 Form 2	2091	760
FI Science Grade 8 Form 1	2395	1083
FI Science Grade 8 Form 2	1833	807
FI Science Grade 11 Form 1	2091	1016
FI Science Grade 11 Form 2	1593	694

Table 2.20

2011–2012 MI–Access Functional Independence—Accessing Print

Percent of Students by Subgroup and Form

	N	Female	Male	Black	Hispanic	White	Economic Disadv	ELL	FLEP
Grade 3									
All Forms	1508	32.8	67.2	24.8	8.0	61.6	74.6	6.1	0.0
Form 1*	518	34.9	65.1	24.1	8.9	60.6	71.2	6.4	0.0
Form 2	294	31.6	68.4	21.4	9.2	64.6	77.9	7.8	0.0
Form 3	327	32.7	67.3	24.8	5.5	63.6	74.9	4.3	0.0
Form 4	369	30.9	69.1	28.5	7.9	58.8	76.4	6.0	0.0
Grade 4									
All Forms	1746	33.6	66.4	25.3	7.0	63.0	74.7	5.8	0.2
Form 1*	588	34.6	65.4	23.0	6.3	65.1	73.1	6.3	0.2
Form 2	335	34.0	66.0	31.0	6.9	57.6	74.0	4.2	0.3
Form 3	417	29.3	70.7	28.8	7.4	60.4	77.2	5.0	0.2
Form 4	406	36.5	63.5	20.4	7.9	67.0	75.4	7.4	0.0
Grade 5									
All Forms	1885	35.2	64.8	27.1	7.0	61.5	72.8	5.3	0.2
Form 1*	691	34.6	65.4	22.8	5.9	68.3	72.3	4.6	0.1
Form 2	432	37.3	62.7	24.1	8.1	60.0	70.8	6.2	0.2
Form 3	389	33.9	66.1	33.4	6.4	57.1	74.3	3.9	0.0
Form 4	373	35.1	64.9	31.9	8.3	55.2	74.5	6.7	0.3
Grade 6									
All Forms	1870	35.3	64.7	27.6	6.6	61.0	72.5	4.9	0.1
Form 1*	639	35.6	64.4	26.3	7.7	60.5	74.1	4.5	0.0
Form 2	435	33.6	66.4	28.5	6.9	59.1	72.6	6.4	0.0
Form 3	399	35.3	64.7	29.8	5.5	58.9	69.2	3.5	0.3
Form 4	397	36.8	63.2	26.4	5.8	65.7	72.8	5.0	0.0
Grade 7									
All Forms	1866	35.3	64.7	29.8	6.2	60.5	72.6	3.4	0.1
Form 1*	709	33.1	66.9	31.9	5.8	59.2	70.2	4.4	0.1
Form 2	379	37.5	62.5	32.5	6.6	55.7	71.8	1.8	0.0
Form 3	376	35.1	64.9	23.9	4.8	68.1	73.7	2.7	0.0
Form 4	402	37.3	62.7	29.4	7.7	60.0	76.9	4.0	0.2
Grade 8									
All Forms	1838	36.5	63.5	29.9	4.5	61.9	72.0	3.4	0.1
Form 1*	670	36.0	64.0	30.2	6.2	60.2	69.4	4.5	0.0
Form 2	436	37.2	62.8	25.7	2.8	68.3	68.6	1.4	0.0
Form 3	372	38.2	61.8	31.5	3.8	59.7	76.3	3.5	0.6
Form 4	360	35.3	64.7	33.1	4.4	59.2	76.4	3.6	0.0
Grade 11									
All Forms	1650	38.4	61.6	25.8	4.1	65.8	67.1	1.9	0.1
Form 1*	682	37.5	62.5	26.7	4.0	66.8	68.2	1.6	0.1
Form 2	366	44.3	55.7	21.9	4.6	69.4	65.8	1.6	0.0
Form 3	287	34.8	65.2	25.8	4.9	62.0	70.0	1.0	0.0
Form 4	315	36.2	63.8	28.6	2.9	62.5	63.2	3.5	0.0

*Form 1 is administered to all students who require the audio, Braille, or enlarged print accommodation. N-counts by accommodation are given in Table 2.12.

Table 2.21
2011–2012 MI–Access Functional Independence—Expressing Ideas
Percent of Students by Subgroup and Form

	N	Female	Male	Black	Hispanic	White	Economic Disadv	ELL	FLEP
Grade 4									
All Forms	1725	33.7	66.3	24.8	7.0	63.5	74.7	5.8	0.2
Form 1*	584	34.6	65.4	23.0	6.0	65.4	73.1	6.0	0.2
Form 2	328	33.8	66.2	29.9	7.0	58.5	74.7	4.3	0.3
Form 3	413	29.5	70.5	28.1	7.5	61.3	77.0	5.1	0.0
Form 4	400	36.8	63.2	20.0	8.0	67.2	75.0	7.5	0.0
Grade 7									
All Forms	1841	35.4	64.6	29.6	6.1	60.7	72.7	3.5	0.2
Form 1*	701	33.0	67.0	31.7	5.7	59.6	70.1	4.4	0.1
Form 2	373	37.5	62.5	31.9	6.4	56.3	71.8	1.9	0.3
Form 3	369	35.2	64.8	23.8	4.9	67.8	73.7	2.7	0.0
Form 4	398	37.7	62.3	29.1	7.8	60.1	77.4	4.0	0.3
Grade 11									
All Forms	1628	38.7	61.3	25.6	4.1	66.0	67.0	1.9	0.1
Form 1*	663	37.7	62.3	26.7	4.1	66.9	68.4	1.5	0.2
Form 2	366	44.3	55.7	21.6	4.6	69.7	66.4	1.6	0.0
Form 3	293	35.8	64.2	25.6	4.8	62.1	68.9	1.4	0.0
Form 4	306	36.6	63.4	28.1	2.9	62.1	62.7	3.6	0.0

*Form 1 is administered to all students who require the audio, Braille, or enlarged print accommodation. N-counts by accommodation are given in Table 2.13.

Table 2.22
2011–2012 MI–Access Functional Independence—Mathematics
Percent of Students by Subgroup and Form

	N	Female	Male	Black	Hispanic	White	Economic Disadv	ELL	FLEP
Grade 3									
All Forms	1353	34.7	65.3	26.4	7.2	60.5	74.9	5.4	0.0
Form 1*	812	34.2	65.8	26.8	6.9	61.3	75.2	5.8	0.0
Form 2	541	35.3	64.7	25.7	7.8	59.1	74.5	4.8	0.0
Grade 4									
All Forms	1551	35.7	64.3	27.0	7.6	61.3	76.0	6.3	0.2
Form 1*	856	36.1	63.9	25.3	7.4	62.7	75.7	5.8	0.2
Form 2	695	35.3	64.7	29.1	7.9	59.6	76.5	6.9	0.1
Grade 5									
All Forms	1721	37.0	63.0	27.7	6.8	61.1	72.1	5.3	0.2
Form 1*	1022	37.1	62.9	27.8	7.0	61.1	71.2	5.1	0.3
Form 2	699	36.8	63.2	27.6	6.6	61.1	73.5	5.7	0.0
Grade 6									
All Forms	1811	36.3	63.7	28.7	6.3	60.5	71.7	4.5	0.1
Form 1*	1034	35.7	64.3	27.2	6.9	60.8	73.1	4.6	0.1
Form 2	777	37.2	62.8	30.6	5.5	60.1	69.9	4.2	0.0
Grade 7									
All Forms	1828	36.8	63.2	31.0	6.0	59.2	73.1	3.2	0.1
Form 1*	1032	37.2	62.8	28.5	6.3	61.8	72.4	3.5	0.1
Form 2	796	36.1	63.9	34.2	5.7	55.9	74.1	2.8	0.1
Grade 8									
All Forms	1831	37.6	62.4	30.6	4.8	61.0	71.8	3.3	0.1
Form 1*	1052	36.8	63.2	29.5	5.4	61.5	72.3	3.3	0.2
Form 2	779	38.9	61.1	32.1	4.0	60.2	71.2	3.2	0.0
Grade 11									
All Forms	1662	38.4	61.6	25.9	4.1	65.7	67.2	1.9	0.1
Form 1*	1008	39.2	60.8	23.7	3.9	67.4	67.7	2.0	0.1
Form 2	654	37.2	62.8	29.5	4.4	62.8	66.4	1.8	0.0

*Form 1 is administered to all students who require the audio, Braille, or enlarged print accommodation. N-counts by accommodation are given in Table 2.14.

Table 2.23
2011–2012 MI–Access Functional Independence—Science
Percent of Students by Subgroup and Form

	N	Female	Male	Black	Hispanic	White	Economic Disadv	ELL	FLEP
Grade 5									
All Forms	1748	36.0	64.0	28.5	6.7	60.4	72.4	4.9	0.2
Form 1*	1001	35.7	64.3	27.1	7.0	61.6	73.3	4.9	0.1
Form 2	747	36.3	63.7	30.5	6.3	58.6	71.2	5.0	0.3
Grade 8									
All Forms	1874	36.8	63.2	30.5	4.7	61.0	71.7	3.3	0.1
Form 1*	1074	37.6	62.4	31.3	4.7	60.8	71.0	3.1	0.2
Form 2	800	35.9	64.1	29.6	4.9	61.3	72.6	3.5	0.0
Grade 11									
All Forms	1661	38.5	61.5	25.8	4.0	65.9	67.2	1.9	0.1
Form 1*	991	38.5	61.5	26.4	4.7	64.2	67.9	2.4	0.1
Form 2	670	38.5	61.5	25.1	3.1	68.1	66.1	1.2	0.0

*Form 1 is administered to all students who require the audio, Braille, or enlarged print accommodation. N-counts by accommodation are given in Table 2.15.

Subgroup Analysis

The raw score mean and standard deviation by content area and grade for the major subgroups are given in the following tables; Table 2.24 for males and females, Table 2.25 for White and Black students and students of other ethnicities, Table 2.26 for economically disadvantaged and non-economically disadvantaged students, and Table 2.27 for accommodated groups. Males had higher average raw scores than the females at all grades of Mathematics and Science. Females had higher average raw scores than males at all grades of Accessing Print and Expressing Ideas. White students had higher raw score averages than the Black students at all grades and content areas. Other students scored in between White and Black students, except in Expressing Ideas grades 4 and 7 and Mathematics grades 4, 8, and 11. In these cases, Other students had higher average raw scores than both White and Black students. Economically disadvantaged students had higher average raw scores than the non-economically disadvantaged students at grades 3 – 7 Mathematics, grade 5 Science, grades 5 and 7 Accessing Print, and grade 7 Expressing Ideas.

Table 2.27 contains the raw score mean and standard deviation for the largest accommodated groups and for those students that did not receive an accommodation. For Expressing Ideas, non-accommodated students had higher average scores than any of the major accommodation groups except students that received the Other accommodations in grade 4 where the two groups had the same average raw scores. For students that received accommodations students that received Scribes had the highest average raw scores in grade 4 and the second highest average raw scores in grades 7 and 11. The ranking of average raw score performance for the rest of the accommodations changed depending on the grade level for Expressing Ideas. For Accessing Print, students with the Reader or Audio CD accommodations had higher average raw scores than students that did not receive an accommodation except at grades 8 and 11 where the non-accommodated students had higher average raw scores than both groups. In some grades, students with Reader did better and in other grades students with the Audio CD did better. Students receiving the Other accommodations did worse than those receiving the Reader and Audio CD except at grade 8 where they scored higher than both groups. For

Mathematics, the results were completely mixed between the accommodated and non-accommodated groups; at some grades the accommodated group had higher average raw scores and at other grades the non-accommodated group had higher average raw scores. For Science, the students receiving the Reader or Audio CD accommodation had higher average raw scores than the non-accommodated group except in grade 11 where non-accommodated students did better than the students that received the Audio CD. The students that received the Other accommodations had higher average raw scores than non-accommodated students in grades 8 and 11 and lower average raw scores in grade 5.

Table 2.24
Raw Score Summary Statistics by Gender*

	Male		Female	
Grade	Mean	SD	Mean	SD
Accessing Print				
3	28.72	7.14	29.03	7.17
4	26.38	7.95	26.67	7.98
5	29.62	7.48	29.78	7.40
6	27.90	7.94	28.00	7.82
7	29.85	7.78	30.62	7.75
8	31.44	7.55	32.37	7.44
11	33.82	6.84	33.88	6.76
Expressing Ideas				
4	2.02	0.77	2.11	0.82
7	2.31	0.88	2.52	0.95
11	2.44	1.00	2.54	1.04
Mathematics				
3	21.45	5.12	20.72	4.88
4	22.12	5.58	21.83	5.50
5	20.49	5.57	19.80	5.44
6	24.35	6.24	22.56	5.89
7	22.48	6.07	21.69	5.45
8	21.73	5.72	20.97	5.63
11	26.69	7.47	24.05	7.50
Science				
5	21.10	5.44	20.31	5.24
8	23.79	6.24	22.99	5.81
11	24.85	6.91	23.17	6.13

*N-counts are given in Tables 2.1 – 2.4.

Table 2.25

Raw Score Summary Statistics by White, Black, and Other Students*

	White		Black		Other	
Grade	Mean	SD	Mean	SD	Mean	SD
Accessing Print						
3	29.66	6.88	26.69	7.58	28.95	6.80
4	27.22	7.78	24.52	8.06	26.75	8.01
5	30.13	7.45	28.76	7.30	29.38	7.66
6	28.70	7.86	26.59	8.03	27.10	7.28
7	30.73	7.78	29.22	7.67	29.78	7.82
8	32.61	7.14	30.14	8.00	31.50	7.57
11	34.73	6.26	31.62	7.57	33.71	7.01
Expressing Ideas						
4	2.06	0.78	2.00	0.80	2.10	0.78
7	2.40	0.91	2.33	0.90	2.41	0.96
11	2.54	0.99	2.36	1.02	2.40	1.12
Mathematics						
3	21.60	4.95	20.17	5.19	21.39	4.91
4	22.42	5.44	21.08	5.65	22.63	5.61
5	20.76	5.35	18.93	5.90	20.65	5.02
6	24.29	6.08	22.34	6.33	24.07	5.78
7	23.01	5.68	20.56	5.82	22.36	5.98
8	22.25	5.64	19.99	5.52	20.90	5.71
11	26.75	7.40	22.58	7.19	26.85	7.65
Science						
5	21.36	5.34	19.69	5.40	20.75	5.10
8	24.45	6.02	21.76	5.79	22.84	6.25
11	25.51	6.57	20.75	5.76	24.64	6.29

*N-counts are given in Tables 2.5 – 2.8.

Table 2.26

Raw Score Summary Statistics by Economically Disadvantaged*

	Economically Disadvantaged		Non-Economically Disadvantaged	
Grade	Mean	SD	Mean	SD
Accessing Print				
3	28.54	7.12	29.65	7.20
4	26.32	7.97	26.96	7.91
5	29.68	7.37	29.65	7.68
6	27.83	7.87	28.21	7.98
7	30.44	7.60	29.53	8.21
8	31.71	7.46	31.95	7.69
11	33.53	7.04	34.49	6.28
Expressing Ideas				
4	2.04	0.79	2.07	0.79
7	2.39	0.89	2.35	0.98
11	2.45	0.99	2.53	1.07
Mathematics				
3	21.22	5.00	21.11	5.18
4	22.36	5.42	21.20	5.86
5	20.46	5.47	19.65	5.64
6	23.84	6.18	23.35	6.16
7	22.41	5.75	21.60	6.10
8	21.41	5.66	21.54	5.80
11	25.60	7.44	25.82	7.89
Science				
5	21.01	5.23	20.30	5.72
8	23.42	6.03	23.69	6.27
11	23.92	6.60	24.79	6.76

*N-counts are given in Tables 2.9 – 2.12.

Table 2.27

Raw Score Summary Statistics by Major Accommodation Group*

	Reader		Audio		Other		Scribe		Non-Accommodated	
Grade	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Accessing Print										
3	29.35	7.01	28.44	6.89	27.68	6.77	NA		27.93	7.52
4	27.30	7.63	27.36	7.51	25.78	7.78	NA		24.79	8.46
5	29.99	7.14	31.26	7.16	28.87	8.13	NA		28.91	7.81
6	28.57	7.40	29.22	7.45	28.50	7.95	NA		26.75	8.38
7	30.54	7.21	30.40	7.55	29.43	9.11	NA		29.93	8.21
8	31.68	7.10	31.10	7.88	32.38	8.27	NA		32.09	7.75
11	33.40	6.55	33.53	6.60	31.65	7.80	NA		34.26	6.94
Expressing Ideas										
4	2.05	0.77	1.94	0.86	2.07	0.79	2.35	0.83	2.07	0.81
7	2.26	0.90	2.36	0.82	2.48	0.91	2.40	0.89	2.50	0.93
11	2.28	1.01	2.52	1.00	1.81	1.33	2.34	0.85	2.59	0.99
Mathematics										
3	21.19	4.86	20.10	5.13	21.23	4.73	NA		21.68	5.52
4	22.15	5.48	22.16	5.23	22.72	5.54	NA		21.74	5.99
5	20.47	5.43	19.76	5.32	20.26	5.42	NA		19.88	5.89
6	24.02	5.94	23.67	6.03	25.08	5.71	NA		23.10	6.51
7	21.92	5.77	22.30	5.17	20.50	5.94	NA		22.43	6.13
8	21.58	5.60	20.60	5.49	22.91	5.43	NA		21.48	5.85
11	25.50	7.84	24.15	7.56	26.89	5.97	NA		25.97	7.43
Science										
5	21.19	5.19	20.42	5.38	19.91	5.93	NA		20.22	5.64
8	23.95	6.07	23.25	6.47	23.33	5.74	NA		23.10	5.98
11	24.45	6.58	23.58	5.97	25.19	6.91	NA		24.10	6.80

*N-counts are given in Tables 2.13 – 2.16.

3. Item Analysis to Facilitate Equating

New secure forms must continually be constructed for future test administrations. The test forms are equated so as to convert the raw scores obtained from two forms of the test so that the scale scores derived from the two forms after conversion will be directly equivalent. Different forms of the test are designed to have comparable item content and similar distributions of item statistics based on field testing. The equating adjusts for unintended differences in difficulty of the forms. The equating also adjusts raw test scores from different forms to a common scale so that identical scale scores earned this year and last year reflect the same level of student achievement, even though the corresponding raw scores may differ.

Equating of the MI–Access Functional Independence Accessing Print, Mathematics, and Science assessments was done using a common item or anchor test design. For Expressing Ideas, the tests consist of a single writing prompt, which is scored out of 4 points. These assessments are not equated and are not used for high-stakes decision making. The description of equating is based on the fall 2010 and fall 2011 forms for grades 3 – 8 and the spring 2011 and spring 2012 forms for grade 11, but applies to all future forms. Anchor items are the same, identical items that appeared in both the 2010-2011 school year form and in the 2011-2012 school year form. For each assessment at each grade, at least 20% of the items were in common between the two forms. The anchor items were used to develop a linking constant that places the Rasch item difficulties from the 2010 school year form on the same logit scale as the 2010-2011 school year form. The linking constant was computed as the difference between the average Rasch difficulty for the anchor items from the 2010-2011 school year form's Winsteps analysis, minus the average Rasch difficulty from the 2011-2012 school year form's Winsteps analysis. For all three content areas, linking constants were computed at each grade the assessment was administered in (i.e., grades 3 – 8 and 11 for Accessing Print and Mathematics and grades 5, 8, and 11 for Science).

Adding this linking constant to the Rasch difficulties for each of the items in the 2011-2012 school year form placed all of the 2011-2012 school year form's Rasch difficulties (and log ability estimates) on the same Rasch logit scale as the 2010-2011 school year form. Then the previous years' linking constants were added to the current year's linking constant to place the 2011-2012 school year form's Rasch log ability scale on the original 2005-2006 scale. Recall that scale scores were developed for each assessment at each grade in the first year by setting the attained cut score to a pre-specified value and the standard deviation to 25. The same linear transformation that was developed in the first year for each assessment at each grade was then applied to the equated Rasch log ability scale for the 2011-2012 school year form to yield equated scale scores.

Since equating involves comparing the Rasch difficulties for the anchor items from the 2011-2012 school year form with those from the 2010-2011 school year form, a plot of those difficulties provides information about the quality of the equating. The plot of the 2011-2012 school year Rasch difficulties versus the 2010-2011 school year Rasch difficulties for the anchor items for each assessment at each grade is given in Appendix A. For Accessing Print, there were 11 anchor or equating items at each grade. For Mathematics, there were 8 equating items at each of grades 3 to 8, and 10 equating items at grade 11. For Science, there were 8 equating items at grade 5, and 10 equating items at grades 8 and 11. Each plot also contains the 45-degree straight line that passes through the mean of the 2011-2012 school year Rasch difficulties and the mean of the 2010-2011 school year Rasch difficulties. The plots show that the Rasch difficulties fall along this 45-degree line as the model requires. Of course, not all points were on or right next to the line due to the inherent error that is in all measurement. Across the 17 assessments, all of the plotted points were close to the identity line. In addition, none of the items had a displacement value greater than the criterion of .5 logits given in the Winsteps manual for potentially deleting an item as an anchor item (Linacre, 2006).

Another way to evaluate the plots is to compute the correlation coefficient between the 2011-2012 Rasch difficulties and the 2010-2011 Rasch difficulties. The correlation coefficient (r) is given in the upper right-hand corner of each plot. Across all seventeen 2011-2012 assessments, the correlations ranged from .870 to .996 with a median correlation of .989. These correlations are as close to 1 as can practically be expected.

Equating involved only the core operational items on each content area and grade level test. Following the equating, the field-test items for each test were calibrated using a fixed anchor test design. For each test, the core items plus the field test items across all forms were calibrated together in a single Winsteps run by fixing or anchoring the core items to the Rasch values obtained during equating. This single run placed all field test items on the same scale as the core operational items.

The Test Characteristic Curve (TCC) and Standard Error Curve (SEC) for each assessment at each grade are given in Appendix B. The cut-scores in each plot are shown with dotted vertical lines. The words “Emerging”, “Attained”, and “Surpassed” provide labels to show which scores were associated with each performance category on the assessment

4. Score Reliability & Summary Statistics

Score reliability was estimated by Cronbach's Coefficient Alpha (Cronbach, 1951) using the item raw score data in R and by the model reliability estimated by using the Rasch model in Winsteps version 3.67.0 (Linacre, 2006). Raw score and scale score summary statistics are also presented in Table 4.1 for Accessing Print, Mathematics, and Science. Data is not reported for Expressing Ideas because the assessment consisted of a single item and hence the reliability was undefined. In addition, scale scores were not computed for Expressing Ideas. Across the grades, the traditional Cronbach's Coefficient Alpha reliability estimates ranged from 0.87 – 0.90 for Accessing Print, 0.80 – 0.88 for Mathematics, and 0.76 – 0.78 for Science. While the Science reliability estimates were lower than those for Accessing Print and Mathematics, they still indicated a reasonably high degree of internal consistency. The model based reliability estimates were somewhat similar to the Cronbach's Coefficient Aphla estimates, but were usually lower than Cronbach's Alpha. The biggest differences between the estimates were for Accessing Print and Mathematics grades 3 – 5.

Table 4.1
Score Reliability and Summary Statistics by Grade

	Grade Level						
Accessing Print	3	4	5	6	7	8	11
Model Reliability	0.83	0.85	0.82	0.84	0.82	0.79	0.78
Cronbach's Alpha	0.87	0.88	0.88	0.88	0.90	0.89	0.90
Raw Score Mean	28.82	26.48	29.67	27.94	30.19	31.78	33.84
Raw Score SD	7.15	7.96	7.45	7.90	7.78	7.52	6.81
Scale Score Mean	2315	2414	2521	2622	2727	2832	3136
Scale Score SD	21.17	22.06	24.04	22.41	22.78	23.83	26.98
	Grade Level						
Mathematics	3	4	5	6	7	8	11
Model Reliability	0.76	0.75	0.78	0.81	0.79	0.79	0.85
Cronbach's Alpha	0.81	0.85	0.83	0.84	0.81	0.80	0.88
Raw Score Mean	21.19	22.08	20.24	23.70	22.19	21.44	25.67
Raw Score SD	5.04	5.55	5.53	6.18	5.86	5.70	7.59
Scale Score Mean	2311	2420	2510	2614	2711	2813	3113
Scale Score SD	21.81	23.01	22.19	20.81	19.11	18.20	24.79
	Grade Level						
Science			5			8	11
Model Reliability			0.75			0.77	0.78
Cronbach's Alpha			0.76			0.78	0.78
Raw Score Mean			20.82			23.49	24.21
Raw Score SD			5.38			6.10	6.67
Scale Score Mean			2500			2799	3108
Scale Score SD			24.86			27.60	22.80

5. Rater Consistency of Expressing Ideas Prompt Scores

The writing prompt responses are scored by human raters. For the core Expressing ideas prompts, 20% of the prompts were scored by two raters at grades 4, 7, and 11, respectively, with 90%, 84%, and 86% exact agreement, respectively. For the field-test Expressing Ideas prompts, 20% were double-scored with between 83% and 87% exact agreement. These agreement rates were similar to rates that have been observed historically for the Expressing Ideas prompts. As seen in Tables 5.1 and 5.2, non-adjacent agreement never occurs for these data.

Table 5.1
Interrater Agreement Rates for Operational Expressing Ideas Prompt Scores

Grade	Total Responses Read	Total Responses Read Twice	Percent of Responses Read Twice	Percent Exact Agreement	Percent Adjacent Agreement	Percent Non-Adjacent Agreement
4	1805	361	20%	90%	10%	0%
7	1912	383	20%	84%	16%	0%
11	1732	346	20%	86%	14%	0%

Table 5.2
Interrater Agreement Rates for Field Test Expressing Ideas Prompt Scores

Grade	Prompt	Total Responses Read	Total Responses Read Twice	Percent of Responses Read Twice	Percent Exact Agreement	Percent Adjacent Agreement	Percent Non-Adjacent Agreement
4	1	965	193	20%	84%	16%	0%
	2	840	168	20%	86%	14%	0%
7	1	1119	224	20%	83%	17%	0%
	2	792	158	20%	85%	15%	0%
11	1	638	128	20%	87%	13%	0%
	2	1086	217	20%	85%	15%	0%

6. Conditional Standard Error of Measurement at Cut-Points

The conditional standard error of measurement was estimated in the raw-score to scale-score conversion tables after equating. These estimates were based on the ratio of raw-score and scale-score standard deviations to scale the conditional SEM associated with each theta as estimated by the Rasch model in Winsteps. Appendix B shows the plot of all conditional standard errors for each assessment in the standard error curve plots. In the plots, the scale score cuts are denoted with a vertical dotted line each graphic. Table 6.1 displays the conditional standard error of measurement at each of the cut points for each of the three content areas.

Table 6.1
Conditional Standard Error of Measurement of Cut-Points by Subject and Grade

	Attained		Surpassed	
Grade	Scale Score	Conditional SEM	Scale Score	Conditional SEM
ELA				
3	2300	7	2315	7
4	2400	6	2415	7
5	2500	6	2511	7
6	2600	6	2614	6
7	2700	6	2713	6
8	2800	6	2820	6
11	3100	6	3129	8
Mathematics				
3	2300	8	2314	9
4	2400	7	2417	8
5	2500	8	2515	9
6	2600	7	2617	8
7	2700	7	2714	8
8	2800	7	2817	7
11	3100	7	3135	10
Science				
5	2500	11	2517	12
8	2800	12	2816	13
11	3100	10	3122	10

7. Classification Accuracy and Consistency

The percent of students at each of the three performance levels by content area and grade are given in Table 7.1. For Accessing Print the largest percent was for *Surpassed* with 45.0% to 72.8%. For Mathematics excluding grades 8 and 11, the largest group was *Surpassed* with 39.5% to 53.5%. For grades 8 and 11, *Attained* had the greatest of students in a performance level category. There were 42.4% of students in the *Attained* category for grade 8 and 50.6% of students in that category for grade 11. For Science, the largest group was *Emerging* with 54.7%, 49.2%, and 38.1% at grades 5, 8, and 11, respectively. The sum of the *Attained* and *Surpassed* percentages in Table 7.1 represents the percentages of students that would be considered proficient for adequate yearly progress (AYP) calculations under NCLB. With the exception of grade 5 Science, more students were proficient on each assessment than not proficient.

Table 7.1
Percent of Students by Proficiency Level

Grade	Emerging		Attained		Surpassed	
	N	%	N	%	N	%
Accessing Print						
3	369	24.5	340	22.5	799	53.0
4	480	27.5	479	27.4	786	45.0
5	358	19.0	257	13.6	1269	67.4
6	283	15.1	414	22.2	1172	62.7
7	185	9.9	322	17.3	1358	72.8
8	138	7.5	439	23.9	1257	68.5
11	141	8.5	522	31.6	987	59.8
Math						
3	379	28.0	361	26.7	613	45.3
4	280	18.1	440	28.4	830	53.5
5	517	30.1	523	30.4	679	39.5
6	468	25.9	538	29.7	803	44.4
7	520	28.5	510	27.9	797	43.6
8	385	21.1	774	42.4	666	36.5
11	519	31.4	837	50.6	298	18.0
Science						
5	954	54.7	333	19.1	457	26.2
8	919	49.2	433	23.2	516	27.6
11	630	38.1	583	35.2	441	26.7

Classification accuracy and consistency are indices of agreement for performance-level classification as a score. Classification accuracy is a way to estimate the difference between true classification and observed classification due to measurement error. Classification consistency is a way to estimate the difference between the observed classification and the classification on a parallel form. The MI–Access FI classification accuracy and consistency indices were calculated using the Rudner based indices in R (Rudner, 2001; 2005; Wyse & Hao, 2012). These indices assume that IRT models are applied on the assessment and that measurement error is normally distributed conditional on ability (Rudner, 2001; Rudner, 2005; Wyse & Hao, 2012). These indices are presented in the following table, Table 7.2. The accuracy indices can be interpreted as the proportion of examinees that would be classified accurately into the performance-level score categories given

infinite replications of identical conditions. The consistency indices can be interpreted as the proportion of examinees that would be classified into the same performance-level score categories on the assessment and a parallel form of the assessment.

Table 7.2
Estimated Classification Accuracy and Consistency by Subject and Grade

	2 Categories Emerging vs Attained plus Surpassed		3 Categories Emerging vs Attained vs Surpassed	
Grade	Accuracy	Consistency	Accuracy	Consistency
ELA				
3	0.92	0.89	0.82	0.75
4	0.92	0.89	0.83	0.76
5	0.94	0.92	0.86	0.80
6	0.94	0.91	0.85	0.79
7	0.95	0.93	0.87	0.83
8	0.96	0.95	0.88	0.83
11	0.97	0.96	0.87	0.81
Mathematics				
3	0.89	0.85	0.76	0.69
4	0.92	0.88	0.80	0.73
5	0.89	0.85	0.77	0.70
6	0.91	0.87	0.79	0.71
7	0.89	0.84	0.78	0.70
8	0.89	0.85	0.77	0.69
11	0.91	0.88	0.85	0.78
Science				
5	0.86	0.81	0.77	0.69
8	0.87	0.82	0.77	0.71
11	0.86	0.81	0.76	0.68

The classification accuracy when categorizing students into the NCLB categories of proficient (*Attained* + *Surpassed*) and not proficient (*Emerging*), ranged from 92 to 97% for Accessing Print, 89% to 92% for Mathematics and 86% to 87% for Science. Classification consistency ranged from 89% to 96% for Accessing Print, 85% to 88% for Mathematics and 81% to 82% for Science. Across all grades and the three content areas, the classification accuracy when categorizing students into three categories (*Emerging*, *Attained*, and *Surpassed*) was 76% to 88% and the classification consistency was 68% to 83%. The lower classification indices were for the tests with the lowest reliability where a three category classification would have the greatest effect on the classification indices. The accuracy indices will be higher than the consistency indices because the former estimates accuracy between observed scores containing measurement error and true scores with no error, whereas the latter estimates consistency between observed scores on parallel forms of

the assessment where both scores contain measurement error.

These estimates represent strong proportions of students classified accurately for an assessment of the length appropriate for students with disabilities such as those that take the MI–Access FI assessments.

For Accessing Print and Mathematics, not only are status level performance categorizations used in school accountability calculations, but information on performance level change on the MI-Access FI assessments is used as a measure of student growth. Performance level change is not computed for Science since the assessments are not given in contiguous grades. Performance level change between last year and this year was computed for grades 4 – 8. Those results are summarized in Table 7.3 for Accessing Print and in Table 7.4 for Mathematics. The first column contains the total number of students assessed in 2011 and matched to 2010 and the second column was that number expressed as a percent. The five subsequent columns show the number and percent of students by performance level change category; significant decline, decline, maintaining, improvement, and significant improvement. Students that obtained an improvement or a significant improvement were counted as proficient due to growth. The tables show that the percentages in the various performance level change categories were quite different across grades. For example, for Accessing Print the percentage of students showing a significant improvement ranged from 4.9% to 19.3%. These percentages show that the pattern of performance level change classifications was not entirely consistent across grades or content areas. In most grades, more students improved and significantly improved than declined or significantly declined. Notable exceptions to this pattern were grade 4 Accessing Print where the improvements and significant improvements were similar to the declines and significant declines and grade 8 Accessing Print, grade 5 Mathematics, and grade 7 Mathematics where more students declined and significantly declined.

Table 7.3
Number and Percent of Students by Performance Level Change Between 2010 and 2011
Accessing Print

Grade	N	Percent Matched	Significant Decline	Decline	Maintaining	Improvement	Significant Improvement
4	1238	70.9	112	332	347	337	110
			9.0	26.8	28.0	27.2	8.9
5	1463	77.7	59	226	312	583	283
			4.0	15.4	21.3	39.8	19.3
6	1449	77.5	101	321	392	458	177
			7.0	22.2	27.1	31.6	12.2
7	1461	78.3	57	262	486	511	145
			3.9	17.9	33.3	35.0	9.9
8	1478	80.6	72	422	573	338	73
			4.9	28.6	38.8	22.9	4.9

Table 7.4

Number and Percent of Students by Performance Level Change Between 2010 and 2011

Mathematics

Grade	N	Percent Matched	Significant Decline	Decline	Maintaining	Improvement	Significant Improvement
4	1086	70.1	64	224	255	424	119
			5.9	20.6	23.5	39.0	11.0
5	1314	76.4	196	524	306	250	38
			14.9	39.9	23.3	19.0	2.9
6	1393	77.0	92	346	311	487	157
			6.6	24.8	22.3	35.0	11.3
7	1412	77.3	97	439	406	390	80
			6.9	31.1	28.8	27.6	5.7
8	1460	80.0	49	380	417	507	107
			3.4	26.0	28.6	34.7	7.0

8. Differential Item Functioning

Differential Item Functioning (DIF) was assessed via the DIFAS computer software (Penfield, 2007). Several DIF statistics were computed with the program, but the ETS category (Zieky, 1993) based on the Mantel Haenszel statistic (Holland & Thayer, 1988) and the Standardized Mean Differences (Dorans & Kulick, 1986) were the ones reported to the Content Advisory Committee (CAC) and the Sensitivity Review Committee (SRC). The ETS scheme categorizes items as follows: A represents negligible DIF, B represents moderate DIF, and C represents large DIF. DIF analyses were run on the following four pairs of comparison groups; gender, white students versus black students, non-economically disadvantaged students versus economically disadvantaged students, and non-accommodated students versus accommodated students. DIF analyses were run on all core and field-test items, but only results for the field test items were shown to the CAC and the SRC.

The number of items by ETS category and grade for each of the four comparison groups is reported in Table 8.1 for Accessing Print, in Table 8.2 for Expressing Ideas, in Table 8.3 for Mathematics, and in Table 8.4 for Science. The number of items favoring each group within each of the four comparison groups is also reported in each table. Table 8.5 shows the percent of items by ETS category across all grades for each content area. Across the grades for Accessing Print, Mathematics, and Science for the gender, race, and economic DIF comparisons, at least 94% of the items were classified with negligible DIF. For Expressing Ideas, race and economically disadvantaged had 100% negligible DIF, while gender had 11.1% negligible DIF and 88.9% moderate DIF. For Accessing Print gender, there was one item classified with large DIF items and for Mathematics gender there was two items classified with large DIF. For Mathematics race/ethnicity, there was one classified as large DIF. Accessing Print economically disadvantaged had also had one item with large DIF. Science had no items classified with large DIF. Similarly, Expressing Ideas had no items with large DIF.

Across all grades for the accommodation comparison, 96.1% of the Mathematics items and 95.5% of the Science items were classified with negligible DIF, and none of the items were classified with large DIF. The results for the accommodation comparison for Accessing Print and Expressing Ideas, however, were noticeably different from all the other results. In these cases, Accessing Print had 79.3% of the items with negligible DIF and Expressing Ideas had 77.8% of the items with negligible DIF. For Accessing Print, there were 11 items classified with large DIF, which is a considerable number of items to be flagged with large DIF. It would appear having an accommodation had a much greater effect on the Accessing Print and Expressing assessments than on the Mathematics and Science assessments. Within Accessing Print, one can see that in general the impact of the accommodation seems to have had a larger impact in earlier grades compared to later grades.

For items classified with moderate or large DIF, the number and percent of such items favoring the reference group (male, white, non-economically disadvantaged, and non-accommodated) and the focal group was determined. Across all grades, the favored group results were mixed by content area and comparison group. In some cases, about the same percentage of items favored each group and in the other cases sometimes a greater percentage of items favored the reference group and sometimes it favored the focal group. The only exception to this pattern was for Expressing Ideas where all of the items with DIF favored females or accommodated students. All field-test items classified with moderate or large DIF were given special attention during the review process by the SRC.

Table 8.1
Accessing Print DIF Summary

	Male vs. Female	White vs. Black	Non-Econ Dis vs. Econ Dis	Non-Accommodated vs. Accommodated
Grade 3				
A: Negligible DIF	78	70	73	43
B: Moderate DIF	1	9	5	30
C: Large DIF	0	0	1	6
Favoring Reference[1]	0	2	5	21
Favoring Focal[1]	1	7	1	15
Grade 4				
A: Negligible DIF	77	80	81	67
B: Moderate DIF	4	1	0	13
C: Large DIF	0	0	0	1
Favoring Reference[1]	0	0	0	8
Favoring Focal[1]	4	1	0	6
Grade 5				
A: Negligible DIF	73	72	78	58
B: Moderate DIF	8	9	3	21
C: Large DIF	0	0	0	2
Favoring Reference[1]	6	1	1	13
Favoring Focal[1]	2	8	2	10
Grade 6				
A: Negligible DIF	81	77	78	65
B: Moderate DIF	0	4	3	16
C: Large DIF	0	0	0	0
Favoring Reference[1]	0	3	1	7
Favoring Focal[1]	0	1	2	9
Grade 7				
A: Negligible DIF	76	75	76	73
B: Moderate DIF	5	6	5	7
C: Large DIF	0	0	0	1
Favoring Reference[1]	4	3	1	6
Favoring Focal[1]	1	3	4	2
Grade 8				
A: Negligible DIF	77	78	81	72
B: Moderate DIF	4	3	0	8
C: Large DIF	0	0	0	1
Favoring Reference[1]	3	2	0	3
Favoring Focal[1]	1	1	0	6
Grade 11				
A: Negligible DIF	70	70	76	70
B: Moderate DIF	10	11	5	11
C: Large DIF	1	0	0	0
Favoring Reference[1]	5	4	3	5
Favoring Focal[1]	6	7	2	6

[1]Number of Category B or C items favoring the Reference or first subgroup given in the column head, e.g. Male, and the number favoring the focal or second subgroup given in the column head, e.g., Female.

Table 8.2

Expressing Ideas DIF Summary

	Male vs. Female	White vs. Black	Non-Econ Dis vs. Econ Dis	Non-Accommodated vs. Accommodated
Grade 4				
A: Negligible DIF	1	3	3	1
B: Moderate DIF	2	0	0	2
C: Large DIF	0	0	0	0
Favoring Reference[1]	0	0	0	0
Favoring Focal[1]	2	0	0	2
Grade 7				
A: Negligible DIF	0	3	3	1
B: Moderate DIF	3	0	0	2
C: Large DIF	0	0	0	0
Favoring Reference[1]	0	0	0	0
Favoring Focal[1]	3	0	0	2
Grade 11				
A: Negligible DIF	0	3	3	3
B: Moderate DIF	3	0	0	0
C: Large DIF	0	0	0	0
Favoring Reference[1]	0	0	0	0
Favoring Focal[1]	3	0	0	0

[1]Number of Category B or C items favoring the Reference or first subgroup given in the column head, e.g. Male, and the number favoring the focal or second subgroup given in the column head, e.g., Female.

Table 8.3
Mathematics DIF Summary

	Male vs. Female	White vs. Black	Non-Econ Dis vs. Econ Dis	Non-Accommodated vs. Accommodated
Grade 3				
A: Negligible DIF	44	42	45	44
B: Moderate DIF	1	4	1	2
C: Large DIF	1	0	0	0
Favoring Reference[1]	1	2	1	1
Favoring Focal[1]	1	2	0	1
Grade 4				
A: Negligible DIF	45	42	44	43
B: Moderate DIF	1	3	2	3
C: Large DIF	0	1	0	0
Favoring Reference[1]	0	4	0	1
Favoring Focal[1]	1	0	2	2
Grade 5				
A: Negligible DIF	42	45	45	45
B: Moderate DIF	4	1	1	1
C: Large DIF	0	0	0	0
Favoring Reference[1]	2	1	0	1
Favoring Focal[1]	2	0	1	0
Grade 6				
A: Negligible DIF	53	51	52	55
B: Moderate DIF	2	4	3	0
C: Large DIF	0	0	0	0
Favoring Reference[1]	0	3	2	0
Favoring Focal[1]	2	1	1	0
Grade 7				
A: Negligible DIF	51	55	52	53
B: Moderate DIF	4	0	3	2
C: Large DIF	0	0	0	0
Favoring Reference[1]	0	0	0	0
Favoring Focal[1]	4	0	3	2
Grade 8				
A: Negligible DIF	52	54	52	55
B: Moderate DIF	3	1	3	0
C: Large DIF	0	0	0	0
Favoring Reference[1]	3	1	1	0
Favoring Focal[1]	0	0	2	0
Grade 11				
A: Negligible DIF	57	55	59	54
B: Moderate DIF	2	5	1	6
C: Large DIF	1	0	0	0
Favoring Reference[1]	1	3	1	3
Favoring Focal[1]	2	2	0	3

[1]Number of Category B or C items favoring the Reference or first subgroup given in the column head, e.g. Male, and the number favoring the focal or second subgroup given in the column head, e.g., Female.

Table 8.4
Science DIF Summary

	Male vs. Female	White vs. Black	Non-Econ Dis vs. Econ Dis	Non-Accommodated vs. Accommodated
Grade 5				
A: Negligible DIF	47	49	51	47
B: Moderate DIF	4	2	0	4
C: Large DIF	0	0	0	0
Favoring Reference[1]	1	2	0	2
Favoring Focal[1]	3	0	0	2
Grade 8				
A: Negligible DIF	59	56	59	58
B: Moderate DIF	1	4	1	2
C: Large DIF	0	0	0	0
Favoring Reference[1]	0	1	0	1
Favoring Focal[1]	1	3	1	1
Grade 11				
A: Negligible DIF	63	62	62	63
B: Moderate DIF	2	3	3	2
C: Large DIF	0	0	0	0
Favoring Reference[1]	1	2	1	0
Favoring Focal[1]	1	1	2	2

[1]Number of Category B or C items favoring the Reference or first subgroup given in the column head, e.g. Male, and the number favoring the focal or second subgroup given in the column head, e.g., Female.

Table 8.5
DIF Summary Across Grades

	Male vs. Female	White vs. Black	Non-Econ Dis vs. Econ Dis	Non-Accommodated vs. Accommodated
Accessing Print				
Total Across All Grades				
A: Negligible DIF	532	522	543	448
B: Moderate DIF	32	43	21	106
C: Large DIF	1	0	1	11
Favoring Reference[1]	14	15	11	63
Favoring Focal[1]	19	28	11	54
Total	565	565	565	565
Percents of Total Across All Grades				
A: Negligible DIF	94.2%	92.4%	96.1%	79.3%
B: Moderate DIF	5.7%	7.6%	3.7%	18.8%
C: Large DIF	0.2%	0.0%	0.2%	1.9%
Favoring Reference[1]	2.5%	5.0%	1.9%	11.2%
Favoring Focal[1]	3.4%	2.7%	1.9%	9.6%
Expressing Ideas				
Total Across All Grades				
A: Negligible DIF	1	9	9	7
B: Moderate DIF	8	0	0	2
C: Large DIF	0	0	0	0
Favoring Reference[1]	0	0	0	0
Favoring Focal[1]	8	0	0	2
Total	9	9	9	9
Percents of Total Across All Grades				
A: Negligible DIF	11.1%	100.0%	100.0%	77.8%
B: Moderate DIF	88.9%	0.0%	0.0%	22.2%
C: Large DIF	0.0%	0.0%	0.0%	0.0%
Favoring Reference[1]	0.0%	0.0%	0.0%	0.0%
Favoring Focal[1]	88.9%	0.0%	0.0%	22.2%

Table 8.5 (cont.)
DIF Summary Across Grades

	Male vs. Female	White vs. Black	Non-Econ Dis vs. Econ Dis	Non-Accommodated vs. Accommodated
Mathematics				
Total Across All Grades				
A: Negligible DIF	344	344	349	349
B: Moderate DIF	17	18	14	14
C: Large DIF	2	1	0	0
Favoring Reference[1]	7	13	5	6
Favoring Focal[1]	12	6	9	8
Total	363	363	363	363
Percents of Total Across All Grades				
A: Negligible DIF	94.8%	94.8%	96.1%	96.1%
B: Moderate DIF	4.7%	5.0%	3.9%	3.9%
C: Large DIF	0.6%	0.3%	0.0%	0.0%
Favoring Reference[1]	1.9%	3.6%	1.4%	1.7%
Favoring Focal[1]	3.3%	1.7%	2.5%	2.2%
Science				
Total Across All Grades				
A: Negligible DIF	169	167	172	168
B: Moderate DIF	7	9	4	8
C: Large DIF	0	0	0	0
Favoring Reference[1]	2	7	1	3
Favoring Focal[1]	5	2	3	5
Total	176	176	176	176
Percents of Total Across All Grades				
A: Negligible DIF	96.0%	94.9%	97.7%	95.5%
B: Moderate DIF	4.0%	5.1%	2.3%	4.5%
C: Large DIF	0.0%	0.0%	0.0%	0.0%
Favoring Reference[1]	1.1%	4.0%	0.6%	1.7%
Favoring Focal[1]	2.8%	1.1%	1.7%	2.8%

[1]Number of Category B or C items favoring the Reference or first subgroup given in the column head, e.g. Male, and the number favoring the focal or second subgroup given in the column head, e.g., Female.

9. Interrelations Among Strands within Measures

One important source of reliability evidence is the consistency of the relations of test subcomponents – interrelations among strands within the test. The correlations were computed based on subscore raw scores and estimated as Pearson product-moment correlations in R.

The correlations between multiple choice (Accessing Print) and constructed response (Expressing Ideas) scores by grade for grades 4, 7, and 11 are given in Table 9.1. The correlations were .30, .41, and .45 at grades 4, 7, and 11, respectively. Table 9.2 contains mean Accessing Print scores for each possible Expressing Ideas score. The results in Table 9.2 suggest that mean Accessing Print scores generally increased as the Expressing Ideas scores increased. The one exception to this was for grade 7 where students that earned a zero had slightly higher mean scores for Accessing Print than students that got a score of 1. Expressing Ideas scores of 0 are difficult to interpret in a consistent way since this score results from a number of alternative non–scorable responses or condition codes. In addition, it is important to note that student's that got a condition codes of D, which was a blank on both prompts, were considered not valid for Expressing Ideas and there scores were not included in the score of zero responses. This is also the reason for the differences in the N counts for Accessing Print and Expressing Ideas. Table 9.3 contains the Accessing Print and Expressing Ideas strand intercorrelations by grade. Across the grades, the three types of passages were typically correlated among each other in the high 0.40s and 0.50s. The passages each correlated very highly with text comprehension, but this was not surprising since the text comprehension score consisted of the three passages. At grades 4, 7, and 11, Expressing Ideas scores correlated in the 0.20s to the low .40s with each of the passage types and with Word Recognition. These were moderately high correlations given the maximum score was only four points for Expressing Ideas. The Word Recognition scores tended to correlate from the 0.30s to the 0.50s with the passages and text comprehension scores. These correlations tended to increase with grade.

Table 9.4 contains the Mathematics strand intercorrelations for grades 3 to 8, and Table 9.5 contains the intercorrelations for grade 11. Across the grades, the intercorrelations among Mathematics strands ranged from 0.11 to 0.65. The lower observed correlations tended to be associated with the strands that had only a couple items.

Table 9.6 contains the Science strand intercorrelations. The correlations between Life Science, Psychical Science and Earth Science tended to be in the 0.40s and 0.50s. The correlations for the Constructing and Reflecting tended to be lower with each other and the other content areas. These lower correlations were a function of the fact that the items that measured Constructing and Reflecting tended to be few in number in comparison to the other content areas. These resulted mirrored some of the correlations for Mathematics that were lower for some of the strands that had fewer items.

The N, mean, standard deviation, and Cronbach's Coefficient Alpha along with the minimum and maximum score of the strand scores were also computed. These summary statistics are given in Table 9.7 for Accessing Print and Expressing Ideas, in Table 9.8 for Mathematics, and in Table 9.9 for Science. Most of the results presented in these tables are not that unexpected. There were differences in each statistic depending on the grade, content area, and strand. Most of the strands have fairly decent Cronbach's Coefficient Alpha estimates, except for the strands that had only a small number of items. These strands tended to have poor Cronbach's Coefficient Alpha estimates. This is not that unexpected as it is well known that Cronbach's Coefficient Alpha is impacted by the number of items included when computing the statistic.

Table 9.1
Correlations between Multiple Choice (Accessing Print) and Constructed Response (Expressing Ideas) Scores

Grade	N	Correlation
4	1724	0.30
7	1840	0.41
11	1620	0.45

Table 9.2
Mean Accessing Print Score by Expressing Ideas Prompt Score

			Accessing Print	
Grade	Expressing Ideas Score	N	Mean	Standard Deviation
4	0	30	18.6	8.3
	1	331	21.7	7.7
	2	951	26.6	7.9
	3	343	29.9	6.9
	4	69	31.5	7.3
7	0	38	24.2	9.5
	1	212	23.4	8.2
	2	821	28.8	7.7
	3	546	32.8	7.0
	4	223	35.0	5.1
11	0	68	24.9	8.5
	1	152	27.9	8.0
	2	614	32.3	7.7
	3	510	35.8	5.8
	4	276	37.9	4.0

Table 9.3

Accessing Print & Expressing Ideas Strand Pearson Product-Moment Intercorrelations by Grade

	Informational Passage	Narrative Passage	Functional Passage	Word Recognition	Text Comprehension
Grade 3					
Narrative Passage	0.55				
Functional Passage	0.52	0.57			
Word Recognition	0.30	0.32	0.33		
Text Comprehension	0.82	0.86	0.83	0.38	
Grade 4					
Narrative Passage	0.52				
Functional Passage	0.49	0.54			
Word Recognition	0.35	0.37	0.46		
Text Comprehension	0.81	0.84	0.82	0.48	
Expressing Ideas	0.21	0.25	0.28	0.31	0.30
Grade 5					
Narrative Passage	0.51				
Functional Passage	0.49	0.52			
Word Recognition	0.36	0.40	0.49		
Text Comprehension	0.81	0.83	0.82	0.51	
Grade 6					
Narrative Passage	0.55				
Functional Passage	0.52	0.55			
Word Recognition	0.42	0.39	0.46		
Text Comprehension	0.84	0.83	0.82	0.51	
Grade 7					
Narrative Passage	0.55				
Functional Passage	0.55	0.52			
Word Recognition	0.48	0.46	0.47		
Text Comprehension	0.85	0.82	0.83	0.57	
Expressing Ideas	0.27	0.27	0.32	0.39	0.34
Grade 8					
Narrative Passage	0.60				
Functional Passage	0.54	0.52			
Word Recognition	0.49	0.47	0.50		
Text Comprehension	0.87	0.84	0.81	0.58	
Grade 11					
Narrative Passage	0.62				
Functional Passage	0.48	0.49			
Word Recognition	0.50	0.50	0.52		
Text Comprehension	0.87	0.86	0.76	0.61	
Expressing Ideas	0.38	0.34	0.31	0.41	0.41

Table 9.4

Mathematics Strand Pearson Product-Moment Intercorrelations for Grades 3 – 8

	Numbers & Operations	Measurement	Geometry	Data & Probability
Grade 3				
Measurement	0.55			
Geometry	0.47	0.53		
Data & Probability	0.42	0.42	0.43	
Grade 4				
Measurement	0.60			
Geometry	0.44	0.38		
Data & Probability	0.41	0.35	0.30	
Grade 5				
Measurement	0.64			
Geometry	0.35	0.32		
Data & Probability	0.46	0.42	0.22	
Grade 6				
Measurement	0.65			
Geometry	0.30	0.31		
Data & Probability	0.42	0.41	0.21	
Grade 7				
Measurement	0.59			
Geometry	0.28	0.30		
Data & Probability	0.48	0.47	0.23	
Grade 8				
Measurement	0.60			
Geometry	0.39	0.42		
Data & Probability	0.41	0.34	0.23	
Algebra	0.23	0.19	0.11	0.17

Table 9.5

Mathematics Strand Pearson Product-Moment Intercorrelations for Grade 11

	Patterns & Relationships	Geometry & Measurement	Data & Probability	Numbers & Operations
Grade 11				
Geometry & Measurement	0.58			
Data & Probability	0.38	0.46		
Numbers & Operations	0.53	0.69	0.40	
Algebra	0.37	0.46	0.30	0.51

Table 9.6

Science Strand Pearson Product-Moment Intercorrelations

	Constructing	Reflecting	Life Science	Physical Science
	Grade 5			
Reflecting	0.13			
Life Science	0.29	0.14		
Physical Science	0.33	0.16	0.51	
Earth Science	0.26	0.13	0.46	0.50
	Grade 8			
Reflecting	0.21			
Life Science	0.33	0.32		
Physical Science	0.37	0.35	0.47	
Earth Science	0.35	0.35	0.48	0.53
	Grade 11			
Reflecting	0.12			
Life Science	0.32	0.27		
Physical Science	0.29	0.27	0.48	
Earth Science	0.29	0.27	0.43	0.46

Table 9.7

Accessing Print & Expressing Ideas Strand Summary Statistics

	N	Minimum Score	Maximum Score	Mean	Standard Deviation	Cronbach's Alpha
Grade 3						
Informational Passage	1508	0	7	4.34	1.74	0.58
Narrative Passage	1508	0	7	4.38	1.93	0.67
Functional Passage	1508	0	7	4.15	1.71	0.58
Word Recognition	1508	0	20	15.98	4.12	0.86
Text Comprehension	1508	0	21	12.87	4.49	0.80
Grade 4						
Informational Passage	1745	0	7	3.61	1.78	0.54
Narrative Passage	1745	0	7	4.39	1.88	0.64
Functional Passage	1745	0	7	4.32	1.85	0.54
Word Recognition	1745	0	20	14.26	4.71	0.86
Text Comprehension	1745	0	21	12.22	4.54	0.80
Expressing Ideas	1724	0	4	2.05	0.79	NA
Grade 5						
Informational Passage	1884	0	7	3.99	1.75	0.52
Narrative Passage	1884	0	7	4.86	1.75	0.62
Functional Passage	1884	0	7	4.77	1.78	0.52
Word Recognition	1884	0	20	16.05	4.27	0.87
Text Comprehension	1884	0	21	13.62	4.32	0.79
Grade 6						
Informational Passage	1869	0	7	4.56	1.87	0.64
Narrative Passage	1869	0	7	5.14	1.70	0.63
Functional Passage	1869	0	7	4.91	1.73	0.64
Word Recognition	1869	0	20	13.33	4.68	0.85
Text Comprehension	1869	0	21	14.61	4.41	0.82
Grade 7						
Informational Passage	1865	0	7	4.87	1.80	0.64
Narrative Passage	1865	0	7	5.42	1.56	0.60
Functional Passage	1865	0	7	5.27	1.66	0.64
Word Recognition	1865	0	20	14.63	4.61	0.87
Text Comprehension	1865	0	21	15.56	4.18	0.82
Expressing Ideas	1840	0	4	2.38	0.91	NA
Grade 8						
Informational Passage	1834	0	7	5.06	1.81	0.67
Narrative Passage	1834	0	7	5.55	1.58	0.64
Functional Passage	1834	0	7	5.55	1.58	0.67
Word Recognition	1834	0	20	15.61	4.29	0.86
Text Comprehension	1834	0	21	16.17	4.17	0.83
Grade 11						
Informational Passage	1650	0	7	5.16	1.77	0.67
Narrative Passage	1650	0	7	5.29	1.67	0.66
Functional Passage	1650	0	7	6.04	1.34	0.67
Word Recognition	1650	0	20	17.36	3.60	0.87
Text Comprehension	1650	0	21	16.49	3.99	0.83
Expressing Ideas	1620	0	4	2.48	1.02	NA

Table 9.8

Mathematics Strand Summary Statistics

	N	Minimum Score	Maximum Score	Mean	Standard Deviation	Cronbach's Alpha
Grade 3						
Numbers & Operations	1353	0	10	6.43	2.03	0.58
Measurement	1353	0	8	5.21	1.93	0.60
Geometry	1353	0	9	7.00	1.64	0.53
Data & Probability	1353	0	3	2.55	0.74	0.48
Grade 4						
Numbers & Operations	1550	0	16	11.35	3.57	0.80
Measurement	1550	0	8	5.73	1.72	0.55
Geometry	1550	0	4	3.40	0.82	0.37
Data & Probability	1550	0	2	1.60	0.64	0.45
Grade 5						
Numbers & Operations	1719	0	16	10.52	3.38	0.74
Measurement	1719	0	10	6.49	2.07	0.57
Geometry	1719	0	2	1.55	0.62	0.20
Data & Probability	1719	0	2	1.68	0.57	0.35
Grade 6						
Numbers & Operations	1809	0	18	11.91	3.73	0.77
Measurement	1809	0	12	7.92	2.36	0.62
Geometry	1809	0	2	1.59	0.58	0.22
Data & Probability	1809	0	3	2.28	0.78	0.40
Grade 7						
Numbers & Operations	1827	0	17	9.99	3.26	0.69
Measurement	1827	0	12	8.08	2.33	0.60
Geometry	1827	0	3	2.16	0.77	0.18
Data & Probability	1827	0	3	1.96	0.94	0.42
Grade 8						
Numbers & Operations	1825	0	17	9.91	3.17	0.65
Measurement	1825	0	10	6.62	1.99	0.59
Geometry	1825	0	3	2.31	0.81	0.34
Data & Probability	1825	0	3	1.84	0.85	0.18
Algebra	1825	0	2	0.76	0.72	0.19
Grade 11						
Patterns & Relationships	1654	0	4	3.03	1.20	0.66
Geometry & Measurement	1654	0	16	10.94	3.14	0.73
Data Analysis & Statistics	1654	0	2	1.49	0.66	0.29
Numbers & Operations	1654	0	15	8.63	3.25	0.72
Algebra	1654	0	3	1.59	1.00	0.43

Table 9.9

Science Strand Summary Statistics

	N	Minimum Score	Maximum Score	Mean	Standard Deviation	Cronbach's Alpha
Grade 5						
Constructing	1744	0	2	1.28	0.68	0.17
Reflecting	1744	0	1	0.87	0.34	NA
Life Science	1744	0	13	7.36	2.22	0.49
Physical Science	1744	0	12	6.85	2.37	0.55
Earth Science	1744	0	7	4.45	1.56	0.47
Grade 8						
Constructing	1868	0	2	1.50	0.64	0.25
Reflecting	1868	0	2	1.26	0.72	0.20
Life Science	1868	0	14	8.20	2.51	0.53
Physical Science	1868	0	14	7.38	2.46	0.49
Earth Science	1868	0	8	5.15	1.73	0.52
Grade 11						
Constructing	1654	0	2	1.17	0.65	0.06
Reflecting	1654	0	2	1.25	0.70	0.11
Life Science	1654	0	14	7.17	2.83	0.62
Physical Science	1654	0	15	7.37	2.53	0.48
Earth Science	1654	0	12	7.24	2.31	0.52

10. Summary of Items Flagged for Difficulty, Discrimination, and Model Fit

The Michigan Department of Education Bureau of Assessment and Accountability established the following criteria for flagging items based on their difficulty, discrimination and model fit:

PL flag if the adjusted p-value is less than 0.33

PH flag if the adjusted p-value is greater than 0.90

CL flag if the item-total correlation is less than 0.25

MH flag if infit mean-square or outfit mean-square is greater than 2.0

MM flag if infit mean-square is between 1.5 – 2.0 and outfit mean-square is less than 2.0 or outfit mean-square is between 1.5 – 2.0 and infit mean-square is less than 2.0

TP flag if infit mean-square is less than 0.5 and outfit mean-square is less than 1.5 or outfit mean-square is less than 0.5 and infit mean-square is less than 1.5.

The adjusted p-value is equal to the item mean divided by the maximum item score. The infit and outfit mean-squares are statistics that provide information on how well items are fitting the Rasch model. Both statistics have an expectation of 1. For infit, values substantially less than 1 indicate dependency in the data; values substantially above 1 indicate noise. For outfit, values substantially less than 1 indicate dependency in the data; values substantially greater than 1 indicate the presence of unexpected outliers. Infit is sensitive to unexpected behavior affecting responses to items near the person's measure level whereas outfit is more sensitive to unexpected behavior by persons on items far from the person's measure level (Linacre, 2006).

Table 10.1 summarizes the number of flagged items by level, content area, and grade. The table also shows the total number of items for each assessment. Across grades and content areas, 0 to 5 items were flagged for the lower bound difficulty (PL) and 0 to 6 items were flagged for the upper bound difficulty (PH) except at grade 11 Accessing Print where 16 items were flagged. Across grades 3 – 8 and grade 11 for Accessing Print, there were 6 to 15 items flagged for discrimination. The results for Mathematics were similar where 5 to 16 items were flagged for discrimination (CL) depending on the grade. For Science, the number of items flagged for discrimination was higher than Accessing Print and Mathematics with 24 to 33 items flagged for discrimination (CL). The higher number of items flagged for Science is a bit worrisome and helps explain in part the lower Cronbach's Coefficient Alpha estimates reported in the Chapter 4 on score reliability and summary statistics and the lower classification accuracy and consistency estimates in Chapter 7. No items were flagged for discrimination (CL) for Expressing Ideas.

Across grades and content areas, only one item at grades 7 and 8 Accessing Print, two items for Expressing Ideas grade 11, and one item at grade 8 Mathematics were flagged for the most severe model fit flag (MH). For the MM model fit flag, 0 to 8 items were flagged across the grades of Accessing Print, 0 or 1 items were flagged for Expressing Ideas, 0 or 1 items were flagged across the grades of Mathematics, and no items were flagged for Science. For the TP model fit flag, no items were flagged for Expressing Ideas, Mathematics, and Science. There was one item flagged in grades 7, 8, and 11 for Accessing Print. The model item fit statistics suggest that the Rasch model fitted fairly well for these data.

Table 10.1
Number of Flagged Items

Grade	Total Number of Items	Difficulty Flag		Item-Total Correlation Flag[3]	Mean-square Fit Flag		
		PL[1]	PH[2]		MH[4]	MM[5]	TP[6]
Accessing Print							
3	79	0	6	13	0	0	0
4	81	3	0	15	0	4	0
5	81	2	3	15	0	2	0
6	81	1	4	10	0	4	0
7	81	1	3	10	1	3	1
8	81	5	1	6	1	1	1
11	81	0	16	6	0	8	1
Expressing Ideas							
4	3	0	0	0	0	0	0
7	3	0	0	0	0	0	0
11	3	0	0	0	2	1	0
Mathematics							
3	46	1	2	16	0	0	0
4	46	1	3	12	0	0	0
5	46	2	2	10	0	1	0
6	55	1	1	5	0	0	0
7	55	2	1	14	0	0	0
8	55	2	2	9	1	0	0
11	60	3	1	8	0	1	0
Science							
5	51	5	0	24	0	0	0
8	60	5	1	31	0	0	0
11	65	5	0	33	0	0	0

[1]PL = p-value < 0.33
[2]PH = p-value > 0.90
[3]Item-total correlation < 0.25
[4]Infit Mean-square or Outfit Mean-square > 2.0
[5]Infit Mean-square between 1.5 – 2.0 and Outfit Mean-square < 2.0
or Outfit Mean-square between 1.5 – 2.0 and Infit Mean-square < 2.0
[6]Infit Mean-square < 0.5 and Outfit Mean-square < 1.5
or Outfit Mean-square < 0.5 and Infit Mean-square < 1.5

11. Spring Online Pilot Analyses

The Michigan Department of Education Bureau of Assessment and Accountability explored the use of an online answer document in the Spring 2012 for grade 11 Mathematics and Science. The online answer document was similar in design and layout to the traditional paper answer document that has been used previously for MI-Access. The answer document was completed by the teacher by transferring the student responses from the test booklet to the online answer document. The procedure for filling out the online document was very similar to the process used with the paper document. In both cases, the teacher transfers the student answers from test booklet directly to the answer document. The student does not fill out their own answer document. Hence, the use of the answer document is not an online pilot in the traditional sense since the student did not receive test questions on the computer and did not input their answers into the computer. Scores for the online answer document were used operationally to assign scores to students since students did not interact with the answer documents in either the paper or online versions. The online pilot did not include Accessing Print or Expressing Ideas since these two separate assessments were included in the same test booklet and the student does write their own response to the Expressing Ideas prompt in the test booklet. A simple transfer strategy for getting these responses into the computer was not possible as part of the pilot.

Table 11.1 shows the demographic breakdowns of all students that took each assessment, the students that had their scores input through the online system, and the students whose scores were recorded on the paper answer document for Functional Independence. The tables indicate that the demographic characteristics were notably different from each other in terms of the genders of the students, percentage of white students, and the percentage of students that were economically disadvantaged. For the online version of the answer document, a greater proportion of students were female, a lower percentage of the students were white, and a greater percentage of the students were economically disadvantaged. Most of the other demographic characteristics for the two populations were similar. The tables also show that more students used the paper answer document compared to the online answer document.

Table 11.1

2011–2012 Percent of Students by Subgroup and Content Area for Online and Paper Forms

	N	Female	Male	Black	Hispanic	White	Economic Disadv
	Mathematics						
All Forms	1654	38.4	61.6	26.0	4.1	65.6	67.2
Online	180	44.4	55.6	26.1	4.4	62.2	71.1
Paper	1474	37.7	62.3	26.0	4.1	66.0	66.7
	Science						
All Forms	1654	38.5	61.5	25.9	4.1	65.8	67.2
Online	173	43.9	56.1	26.0	4.0	62.4	71.1
Paper	1481	37.9	62.1	25.9	4.1	66.2	66.7

Table 11.2 provides scale score summaries for the students that had their scores input on online and on paper. For Mathematics the online students had average scale scores that were one point higher and for science the average scale scores were the same. The standard deviations were slightly higher for paper students for both content areas. The similarity in scores from the two methods of recording scores is desirable. However, it is important to note that having one's scores recorded on paper or online was not random. Differences or similarities in scores for the two populations in the tables could be a function of some of these similarities and differences between the populations or the way that the student scores were recorded.

Table 11.2
2011–2012 Scale Score Summaries by Online and Paper –

	Online			Paper		
Content Area	Mean	SD	N	Mean	SD	N
Mathematics	3113	24.05	180	3112	24.89	1474
Science	3108	21.23	173	3108	22.89	1481

Table 11.3 provides some DIF comparisons for the online and paper answer documents for the operational items that were used to assign scores to students. Ideally, the amount of DIF for the paper and online versions of the answer document should show very little to no DIF. For Mathematics, only one item exhibited DIF and for Science no items exhibited DIF. This is what one would hope for in terms of the functioning of the items for the online and paper answer documents.

Table 11.3
Online and Paper DIF Summary for Operational Items

	Paper vs. Online N	Paper vs. Online %
Mathematics		
A: Negligible DIF	39	97.5%
B: Moderate DIF	1	2.5%
C: Large DIF	0	0
Favoring Reference[1]	1	2.5%
Favoring Focal[1]	0	0.00%
Science		
A: Negligible DIF	45	100.0%
B: Moderate DIF	0	0.0%
C: Large DIF	0	0.0%
Favoring Reference[1]	0	0.0%
Favoring Focal[1]	0	0.0%

[1]Number of Category B or C items favoring the Reference or first subgroup given in the column head, e.g. Paper, and the number favoring the focal or second subgroup given in the column head, e.g., Online.

12. Verification of Psychometric Procedures

Assessment and Evaluation Services (AES) served as the subcontractor for the Independent Psychometric Quality Assurance Review. AES reviewed and replicated all psychometric procedures connected to the item analysis, scaling and equating of the assessments. AES has experience in performing quality control services in testing programs in Ohio, New Jersey, New York, Virginia, and Washington. In those states, AES verifies similar psychometric analyses as described in the Michigan program. AES has verified analyses involving Item Response Theory (IRT) equating, scaling, and item analysis. AES staff has expertise in IRT models including, but not limited to, the Rasch model, Partial Credit Model, and three-parameter IRT model. AES also has experience in coordinating this work with various contractors.

In the past, AES has found that working with the psychometric staff of another company and/or state department requires extensive planning and coordinated scheduling. It is essential that the quality control work be extensive and accurate, but it is equally important that it be completed in a timely fashion so that overall project schedules can be met. This requires that both AES and the psychometric staff of the prime contractor and/or state department work closely in planning for the transfer of data and analysis results to AES. Likewise, it is important for AES to complete the checking and transfer our results to the state office for verification.

AES met with MI-Access psychometric staff and Questar staff to discuss the plans and schedules regarding the implementation of this contract. The intent was to coordinate the activities between the contractor, state, and AES to ensure that the verification procedures were implemented in a smooth and accurate manner.

AES provided the verification of item analysis scaling and equating activities for the MI-Access 2011-2012 analyses. Verifications were done for post-equating, and field test item analysis. The primary scaling of the ELA, Mathematics, and Science was done with the Rasch model using the partial credit model for open-ended items.

This section provides a description of the steps AES undertook to provide replication of the MI-Access 2011-2012 analyses for the FI assessments. The workflow was organized so that Michigan department psychometric staff and AES staff worked independently on each step. Once major portions of the analysis were completed, AES compiled the two sets of results into a comparison spreadsheet. These spreadsheets were then examined by Michigan department psychometric staff and AES staff to determine if the replication was successful. When discrepancies between department and AES results occurred during the steps they were often resolved before the comparison spreadsheets were completed.

The MI-Access grades 3-8 and grade 11 FI assessments are given in three subject areas: ELA, Mathematics, and Science. ELA and Mathematics were assessed in fall 2011 at grades 3 through 8. Science was assessed at grades 5 and 8. ELA, Mathematics, and Science were assessed in spring 2012 at grade 11. In most cases, the test forms were structured so that forms contained the same census test items and different field test items.

The major analyses for the project are detailed below in steps from the AES perspective. The project has been partitioned into two analysis sets for description; a post assessment file and a field test file. The analyses occurred sequentially. Grades 3-8 analyses were done in the fall and winter of 2011-2012, while the grade 11 analyses were done in the spring and summer of 2012.

Analysis Set 1- Post Equating Scaling for English Language Arts, Mathematics, and Science

The post-equating analysis took place after student work had been scored and was based on a data set which included almost all students. The post-equating analysis checked for the stability of the item difficulty parameters. Differences from the pre-equated values from the bank and the post-equated values from the assessment were examined.

Step 1-Check Data File for Unreasonable Values

In large data files, there are often implausible item response values that are found. This is particularly true in scored data files. AES examined the item score fields for values which were not plausible given the form designation and the item key. When values were found, AES notified Michigan psychometric staff so these instances could be investigated.

Step 2-Initial WINSTEPS Run

The data was analyzed by the WINSTEPS program to develop initial unanchored Rasch difficulties.

Step 3-Develop an Initial Comparison Spreadsheet

The initial Rasch difficulties, n-counts, p-values, and point-biserials were compared. Again very few and very small differences were found.

Step 4- Determine the Equating Constant

The initial Rasch difficulties from Step 3 were compared to item difficulties from the item bank. An equating constant was derived by determining the difference between the means of the two sets of values.

Step 5- Post Equating WINSTEPS Run

An equating constant between bank values and initial WINSTEPS values was developed using common items. The equating constant was then applied to the initial WINSTEPS run to put the item difficulties on the original scale. New raw score to scale score tables were generated.

Step 6- Develop Post Equating Comparison Spreadsheet

Raw-to-scale score tables from Michigan psychometric staff and AES were compared to determine if the scale scores, error term, and performance level were equivalent. The actual differences were very small. A few Rasch thetas were different by .0001, one scale score rounded differently due to a .0001 difference, while all other scale scores, scaled error, and performance levels were identical. The values from Michigan psychometric staff were used when small differences were found.

AES Verification Files for Analysis Set 1

The steps detailed above yield four AES verification files. A verification file consists of three spreadsheets and is used to evaluate the verification analyses.

Primary Spreadsheet - This is provided by the primary technical analyses. In the case of MI-Access, the Michigan psychometric staff performs the primary technical analyses. This spreadsheet provides values for variables for each item or each score point depending on the comparison.

AES Spreadsheet-This is provided by AES. It takes the identical form to the Primary Spreadsheet and contains the AES values for each variable or each score point depending on the comparison.

Verification Spreadsheet-This spreadsheet is simply a comparison of the Primary and AES spreadsheets to see if any differences exist. With numerical variables it is usually a simple subtraction of the Primary value minus the AES value. When alpha numeric codes or flag values are compared differences are noted by the display of the values.

The four AES verification files for the post-equating of MI-Access are focused on checking for data equivalency and verification of equating and scaling results.

Verification File One- Total Classical Statistic Verification
This file is based on the initial item analysis and provides a row in the spreadsheets for each item. The key variables compared were n-counts, p-values, item discrimination indices, and distractor percentages.

Verification File Two- Rasch Item Scaling Verification
This file is based on the WINSTEPS analysis and provides a row in the spreadsheets for each item. The key variables compared were Rasch item difficulty, parameter estimation error, and item fit to the model.

Verification File Three- Raw to Scale Tables
This file is based on the Rasch scaling and provides a row in the spreadsheets for each score point. The key variables compared were the Rasch ability estimates, error of the ability estimate, the derived scale scores, the scale score standard errors, and the performance levels associated with each scale score.

Verification File Four- Operational Assessment Frequency Distribution Table
This file is based on the raw-to-scale score table and provides a row in the spreadsheets for each score point. The key variables compared were the score frequencies, cumulative frequencies, and percentiles.

Analysis Set 2-Field Test Item Analysis

Field test items were analyzed to provide item data for committee review and parameter values for future form scaling and equating. All subject forms contained field test items. The analysis consisted of information about the performance of the item for the total population and item performance for race/ethnicity, gender, accommodated, and economically disadvantaged groups which yielded differential item functioning statistics.

Step 1-Check Data File for Unreasonable Values

In large data files, there are often implausible item response values that are found. This is particularly true in scored data files. AES examined the item score fields for values which were not plausible given the form designation and the item key. When values were found, AES notified Michigan psychometric staff so these instances could be investigated.

Step 2-Item Statistics for the Total Group

Analyses were run by grade/subject on all forms to develop Rasch item parameters, p-values, and point-biserial correlations. The census items were used as base values and their difficulty parameters were fixed so that field test item parameters were placed on the same scale. This method was also used for the Expressing Ideas field test items prompts.

Step 3-Item Statistics for Race/Ethnicity, Gender, Accommodated, and Economically Disadvantaged Groups and calculation of Differential Performance Indicators

The data was analyzed by race/ethnicity, gender, accommodated, and economically disadvantaged groupings. The specific groups investigated were Black/White, male/female, accommodated/non-accommodated, and economically disadvantaged/non-economically disadvantaged. N-counts, p-values, and point-biserials correlations were calculated for each group. Differential item functioning (DIF) statistics developed were the Mantel-Haenszel statistics, the Standardized Mean Differences, and the ETS classifications.

AES Verification Files for Analysis Set 2

The steps detailed above yield four AES verification files. A verification file consists of three spreadsheets and is used to evaluate the verification analysis.

Primary Spreadsheet - This is provided by the primary technical analyses. In the case of MI-Access, the Michigan psychometric staff performs the primary technical analyses. This spreadsheet provides values for variables for each item or each score point depending on the comparison.

AES Spreadsheet-This is provided by AES. It takes the identical form to the Primary Spreadsheet and contains the AES values for each variable or each score point depending on the comparison.

Verification Spreadsheet-This spreadsheet is simply a comparison of the Primary and AES spreadsheets to see if any differences exist. With numerical variables it is usually a simple subtraction of the Primary value minus the AES value. When alpha numeric codes or flag values are compared differences are noted by the display of the values.

The four AES verification files for the field test Analyses of MI-Access are focused on checking for data equivalency, verification or equating and scaling results, and the equivalency or DIF indices.

Verification File One- Total Classical Statistic Verification
This file is based on the item analysis for the total group of students and provides a row in the spreadsheets for each item. The key variables compared were n-counts, p-values, item discrimination indices, and distractor percentages.

Verification File Two- Rasch Item Scaling Verification
This file is based on the WINSTEPS analysis and provides a row in the spreadsheets for each item. The key variables compared were Rasch item difficulty, parameter estimation error, and item fit to the model.

Verification File Three- Sub-group Classical Statistic Verification
This file is based on the item analysis of all items and provides a row in the spreadsheets for each item. The key variables compared were n-counts, p-values, item discrimination indices, and distractor percentages. Rather than one total verification file for each test, a verification file was produced for each sub-group under investigation. Subgroup variables include race/ethnicity, gender, accommodated or economically disadvantaged grouping classifications.

Verification File Four- Differential Item Functioning Analyses
This file is based on analysis of DIF using Mantel- Haenszel and Standardized Mean Difference indices to detect differential item functioning and provides a row in the spreadsheets for each item. The key variables compared for each race/ethnicity, gender, accommodated or economically disadvantaged groupings are the Chi Square, Delta, Standardized Mean Difference, and the ETS classifications.

13. Validity Argument

An important part of ensuring the technical quality of assessments is examining the validity of the score interpretations and uses of the assessments. The approach taken in this chapter regarding validity is similar to the approach presented in Kane (2006). Kane's (2006) view of validity is to view validity as argument where positive and negative evidence is collected and an argument is made based on the evidence for or against the intended uses and interpretations. In this case, the scores from MI-Access FI assessments are used to make decisions about whether or not students' possess adequate mastery of extended grade level content standards by comparing test performance on the assessments to previously established cut-scores. Students that obtain scores that exceeded the cut-scores for the *Attained* performance level are considered to be proficient for that specific grade and assessment. For Accessing Print and Mathematics in grades 4 through 8, previous performance on last year's Michigan assessments is also compared to current achievement to determine if students made enough growth to be considered proficient due to growth. These data are then used in combination with data on other Michigan assessments to make adequate yearly progress determinations for schools. Score results are also presented to districts, schools, teachers, and parents to provide information about how the student performed on the assessments. Data presented in the other chapters of this technical report and from previous years and other special studies for the MI-Access FI assessments constitute the currently complied validity evidence for or against the intended uses and score interpretations. Ideally, the goal is that the positive evidence would outweigh the negative evidence and that the intended uses and interpretations of the scores are more or less supported by the evidence that has been collected.

One important piece of validity evidence relates to the content of the assessment and ensuring that the tests that are given measure an appropriate sampling of the content. In Chapter 1 of this technical report addendum, the test blueprints for the MI-Access FI assessments were provided and outlined. These test blueprints were virtually unchanged from previous test administrations and covered a range of content standards represented in the extended grade level content standards underlying the alternate assessments. In the past, more detailed alignment studies of the content standards for each assessment have been conducted using the Webb alignment procedure (Webb, 1997; 2007) and the Linking for Academic for Learning procedure (Flowers, Wakeman, & Browder, 2009). The results from these studies by and large provided support to the fact that assessments had appropriate alignment, although a few areas for improvement were suggested. These included looking at the specific content of some of the items and ensuring that the items are better written to specific extended grade level content expectations. Additional training has been included in item writer training and item review to address these concerns. This has included additional information on Webb's depth of knowledge and more focused reviews of items, the content standards associated with them, and greater attention being placed on looking at the depth of knowledge of the items. These data provide some positive evidence in support of the uses of the assessments.

Another important piece of validity evidence is related to ensuring that the different forms of the assessment are taken by appropriate samples of students and that the scores from the assessments are comparable for different subgroups of the population. To this end, the Michigan Department of Education has developed a sampling plan to make sure that the forms are distributed in an appropriate fashion to schools and that the samples of students that take in each form have similar characteristics. The results in Chapter 2 provide detailed subgroup analyses and breakdowns of the students that took each form. These results showed that for the most part the groups of students that took each form had fairly similar characteristics. This is what would be expected if the sampling plan was working and being implemented effectively.

Analyses did show that there were some differences in the usage of accommodations across forms and grade levels. These differences across forms did seem to relate to some of the differential item functioning (DIF) analyses that were presented in Chapter 8. In particular, the use of the accommodations for the Audio CD and read aloud accommodations for Accessing Print were higher in the earlier grades. This also corresponded to a situation in which there was greater amount of accommodated versus non-accommodated DIF on the

assessment. It appears that there was an undesirable interaction between accommodation usage and DIF for the Accessing Print assessments. Additional investigations of these data indicated that items that favored accommodation students tended to be on the second part of the assessment with the reading passages and items that favored non-accommodated students tended to be from the first section of the assessment containing the word recognition items. This is a function of how the accommodations are used on each part of these assessments. Students can get the reading passages and answer choices read aloud for the reading passages, but only the stem can be read aloud for the word recognition questions; the answer choices cannot be read aloud. The interactions of test accommodations with performance on the other assessments were less pronounced. The DIF for the accommodated and non-accommodated students provides some negative evidence for how the accommodations for the Accessing Print tests interacted with test content.

In terms of the differences in performance across subgroups that were reported on in Chapter 2, there were some minor differences in subgroup performance at different grade levels for different assessments. In most cases, the differences in performance were less than one or two raw score points, which represented relatively small differences on the assessments. The goal is that these differences would be relatively small and that different groups of students have similar opportunities to learn the content and demonstrate their knowledge through the assessments. The small differences observed provide some positive evidence in support of the uses of the assessments.

Additional evidence on score comparability was presented in Chapter 8 on the DIF analyses and Chapter 11 for analyses comparing the paper answer document to the online answer document for grade 11, Chapter 3 on item analysis to facilitate equating, and Appendix A. The DIF chapter showed that except for the accommodated versus non-accommodated DIF on the Accessing Print assessments and male versus female DIF and accommodated versus non-accommodated DIF on the Expressing Ideas assessments the number of items that showed moderate or large DIF on each assessment was about what one would expect; around 5% of the items showed DIF. In addition, when items did show DIF they did not universally favor one group of students over another. This provides some positive evidence supporting the scores from the assessments.

For the Accessing Print and Expressing Ideas assessments for the three comparisons mentioned above, the amount of DIF was higher than anticipated. This suggests that some additional attention may be needed in regards to these particular subgroups. The results for the accommodations for Accessing Print and Expressing Ideas may be more of signal to the fact some of these students greatly struggle with reading and writing, have disabilities that impact their performance in these areas, and have a need for reading and writing support through which they are provided appropriate accommodations. The challenges that students in alternate assessments have with reading on assessments are well documented (Thurlow, 2010). In-depth consideration of the some of these comparisons and the reasons for the potential differences has been considered at the data review meetings for the test items. Discussions at these meetings with educators from across the state have highlighted the reading challenges that many of the students in these populations have and how these often interact with some of the students' disabilities. The discussions have also affirmed that despite the amount of DIF observed that the teachers feel that the accommodations allowed for the assessments are appropriate and needed.

Chapter 11 on the online answer document shows comparisons and DIF analyses for the use of the online answer document versus the paper answer document in the grade 11 pilots for FI Mathematics and Science. This pilot was not a student facing pilot and consisted primary of teachers either filling out the student answers on a paper answer document or via an online system that mirrored the paper answer documents. Analyses suggested that the scores for the test versions on average were only one scale score point different for Mathematics and no scale score points different for Science. There was only one operational item flagged for DIF for Mathematics and no items flagged for DIF for Science. These results seemed to provide positive evidence that the online and paper answer documents seemed to be functioning in a similar manner in the pilot and that scores can be treated comparably.

Another important piece of validity evidence is making sure that the scores for different test forms that differ in difficulty use appropriate equating designs and adjustments so that students that have the same level of achievement receive the same scores on different forms and over time. Chapter 3 on equating explains the designs used to make sure that each year scores can be placed on a common scale so that scores possess a similar meaning over time. The equating this year followed the same design as previous years and resulted in scores that were placed onto this common scale. Analyses of the equating items showed that none of the items exhibited a large amount of drift over time that would contaminate the linking constants used to produce a common scale. In addition, plots of the Rasch item difficulty parameters in Appendix A showed that the item parameters estimates for the common items used to perform the equating were highly correlated with each other. There was no evidence to suggest that the equating did not perform as intended and that scores were not able to be placed onto a common scale. Also, described in Chapter 12 are all of the steps for performing and replicating the equating and DIF analyses that were done by an independent contractor. All of the equating and DIF results were replicated by the independent contractor, AES. This again provides evidence in support of the scores from the assessments.

Another piece of key evidence for the assessments was data on reliability and on the classification accuracy and consistency of the assessments. These data are found in Chapters 4, 5, 6, 7, and 10. The results from these chapters in most cases are about what one would hope for on the MI-Access FI assessments. Most of the tests have estimated internal consistency reliability estimates that exceeded 0.80 in magnitude, which is an often suggested arbitrary threshold for making sure that the test has appropriate precision. The Science assessments do have some lower reliability estimates than 0.80. This was a function of the greater amount of items that have lower item to total test correlations as is shown by the large amount of items having low item total test correlations in Chapter 10. This suggests that from a technical standpoint the reliability evidence for Science was not as strong as it was for Mathematics or Accessing Print. The reliability still by most standards would be viewed as acceptable. Future assessments administrations should look to try and reduce the number of items with lower item to total test correlations on the Science assessments. This would help to increase the internal consistency reliability estimates.

For the Expressing Ideas assessments, analyses in Chapter 5 indicated that there was sufficient rater consistency of the writing prompts as evidenced by levels of percent agreement by raters that double-scored a sample of the writing prompts. These agreement rates were in the 80s and 90s. In addition, none of the ratings given on the papers that were doubled-scored were non-adjacent scores. Again, these results are about what one would hope for in terms of rater consistency on the Expressing Ideas prompts and provide evidence in support of the uses of the scores from the assessments.

The results on the conditional standard error of measurement are presented in Chapter 6 and test characteristic curves and full conditional standard error of measurement curves are presented in Appendix B. These results show that the conditional standard error of measurement tended to be acceptably small at the cut-scores used to make decisions. The conditional standard error of measurement tended to be higher at the extremes of the score scale where there was less information. Extreme scores tended to have more measurement error than scores that were in the middle of the score scale. The data on model fit in Chapter 10 showed that most of the items displayed adequate fit to the Rasch models used to create the scale scores and perform the equating. Again, the evidence presented does not suggest that there would be issues in the uses of the scores.

Chapter 7 on classification accuracy and consistency provides some evidence about the reliability of the classification decisions that were made on these assessments. The chapter shows that there were different performance level classifications based on the content area and grade level under consideration. These differences are common in many large-scale assessments. The performance levels on the assessments were established several years ago through the application the Bookmark standard setting procedure using representative stakeholders from across Michigan (Lewis, Mitzel, & Green, 1996; Mitzel, Lewis, Patz, & Green, 2001). The results from the standard setting meeting were reviewed and approved by the State Board of

Education and were used to establish the cut-scores for classifying students into different performance categories. The classification decisions had high classification accuracy and consistency, especially when using the scores to make classification decisions used in adequate yearly progress designations. The classification accuracy and consistency was less with three performance levels as compared to two performance levels. This suggests that classification decisions were more reliable when making the decisions for NCLB adequate yearly progress than for other more finely grained decisions.

Chapter 7 also provides data on the growth models used for determining whether the student made enough progress to be considered proficient based on growth. The model used by the state of the Michigan is based on a transition table model in which the performance level categories are broken down into mini-categories to track student performance in Accessing Print and Mathematics from year to year (Martineau, 2007; Wyse, Zeng, & Martineau, 2011). The results suggest that there were some differences in terms of the patterns of growth across grades. In some cases, more students made declines and significant declines and in other cases students made more improves and significant improves. This is a function of where the cut-scores were placed on the original assessments and the varying levels of stringency in terms of meeting these cut-scores that existed across grades. The differential growth patterns across grades, while common, is not preferred because it means that there is the potential for the proficiency due to growth to change across grades. Ideally, the patterns would be consistent from grade to grade. Despite the fact that there were differences and that this is less ideal, the impact on adequate yearly progress calculations should be minimal. The reason for this is that most of the students in each of the grades were already proficient due to status. The only students for which proficiency due to growth comes into play were students who were in the *Emerging mid* or *Emerging high* categories this year and were in the *Emerging low* or *Emerging low* or *mid* categories in the previous years. The rest of the students would already be considered proficient due to status. This is a very small segment of the population.

One last piece of evidence is related to the interrelationships between strands and the reliability of these content strand subscores that was presented in Chapter 9. This chapter suggested that for the most part there were somewhat similar patterns of relationship between the content strands across grade levels. For example, the correlations between the sections of the Accessing Print assessments remained fairly similar across grades. The correlations in Chapter 9 also suggest that the correlations between strands were low to moderate. This suggests that there were differences in performance across strands. This is to be expected if the content strands are measuring somewhat unique aspects of achievement associated with the content area. It is important to point out that even though these correlations were low to moderate that most of the items exhibited acceptable fit to the Rasch model as measured by the INFIT and OUTFIT statistics and shown in Chapter 10. This suggests that the items as a whole seemed to do an acceptable job of measuring the same underlying achievement construct. There was no evidence to suggest that the assessments were functioning in an undesirable way based on the interrelationships between strands.

The results on the reliability of the strand scores were notably lower than reliabilities for the assessments as whole. This suggests that some caution is needed when interpreting the total scores for the strands in various score reports. These scores are not particularly reliable and may change over time. This is especially the case for a few of the content strands with very few test items. These strands tended to have the lowest strand score reliabilities and possessed the greatest lack of precision. However, it is important to point out that these findings are not that different from what is observed in many other large-scale assessments. In many cases, the strand scores are not as reliable as the overall score for the whole assessment. Again, there was no evidence to suggest that the assessments were functioning in an inappropriate manner.

In viewing all of the evidence as a whole, one can see that by and large the positive evidence for the assessments appears to outweigh the negative evidence for the assessments. This suggests that based on the evidence collected that the intended uses and interpretations of the scores seems to be more supported than rejected.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Research Association.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Dorans, N., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, *23*, 355-368.

Flowers, C., Wakeman, S., & Browder, D. (2009). Links for academic learning (LAL): A conceptual model for investigating alignment of alternate assessments based on alternate achievement standards. *Educational Measurement: Issues and Practice, 28*, 25-37.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement (4th edition*, pp. 17-64). Washington, DC: American Council on Education.

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996, June). Standard setting: A bookmark approach. In D. R. Green (Chair), *IRT-based standard setting procedures utilizing behavioral anchoring.* Paper presented at the symposium of the Council of Chief State School Offices National Council of Large-Scale Assessment. Phoenix, AZ.

Linacre, J. M. (2006). *WINSTEPS Rasch measurement computer program*. Chicago: Winsteps.com

Martineau, J. A. (2007, March). *Designing a valid and transparent progress-based value-added accountability model.* Paper presented at the Annual Conference of the American Educational Research Association (AERA), Chicago, IL.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249-281). Mahwah, NJ: Lawrence Erlbaum.

Penfield, R. D. (2007). *DIFAS 4.0 Differential item functioning analysis system*. Thousand Oaks, CA: Sage Publications.

Rudner, L. M. (2001). Computing the expected proportion of misclassified examinees. *Practical Assessment, Research & Evaluation*, *7(14)*. Retrieved from http://PAREonline.net/getvn.asp?v=7&n=14.

Rudner, L. M. (2005). Expected classification accuracy. *Practical Assessment, Research & Evaluation, 10(13)*. Retrieved from http://PAREonline.net/pdf/v10n13.pdf.

Thurlow, M. L. (2010). Steps toward creating fully accessible reading assessments. *Applied Measurement in Education*, *23*, 121-131.

Webb, N. L. (1997). *Criteria for the alignment of expectations and assessments in mathematics and science* (Council of Chief State School Officers and National Institute of Science Education Research Monograph No. 6). Madison: University of Wisconsin-Madison, Wisconsin Center for Educational Research.

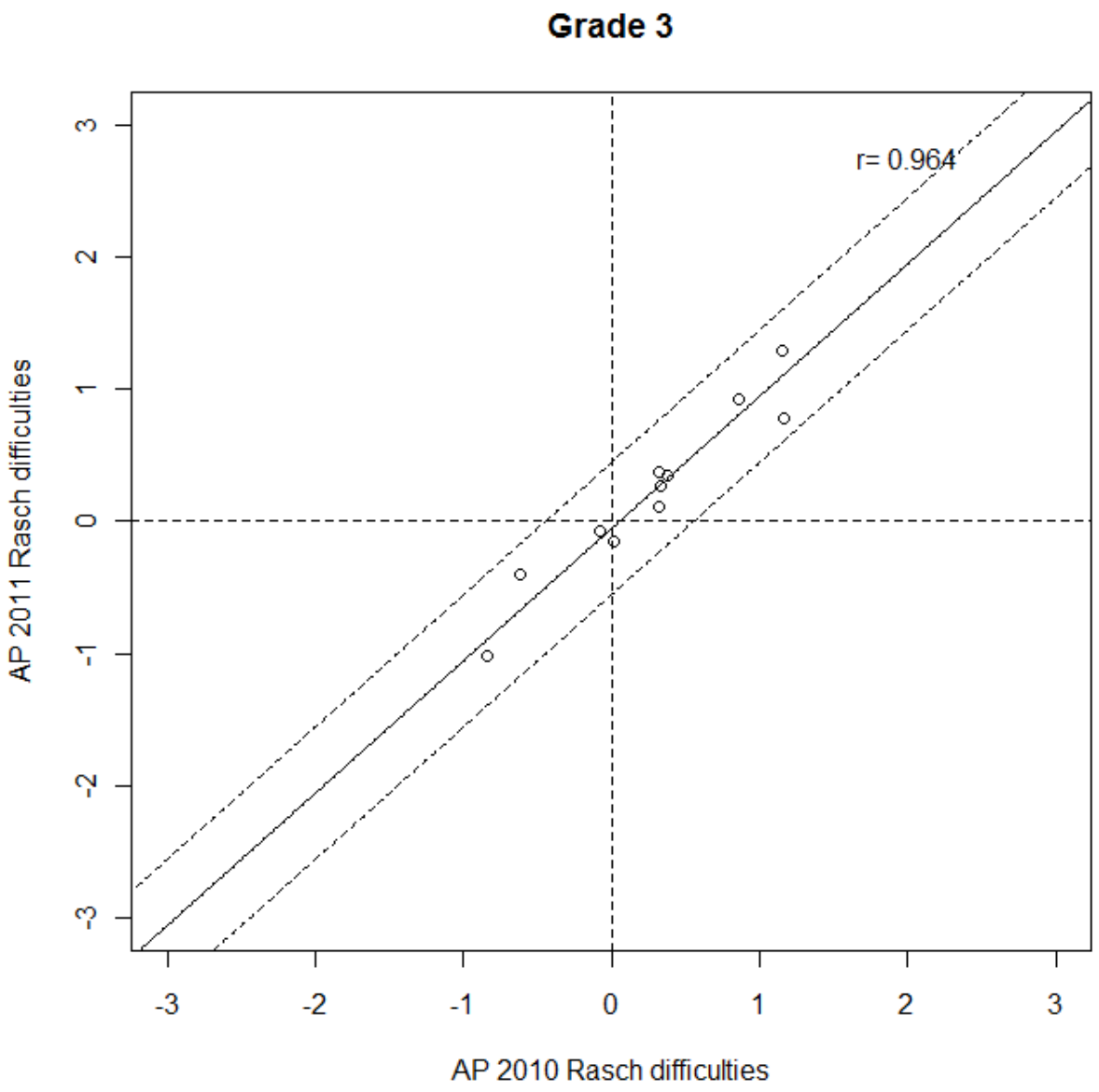
Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, *20*, 7-25.

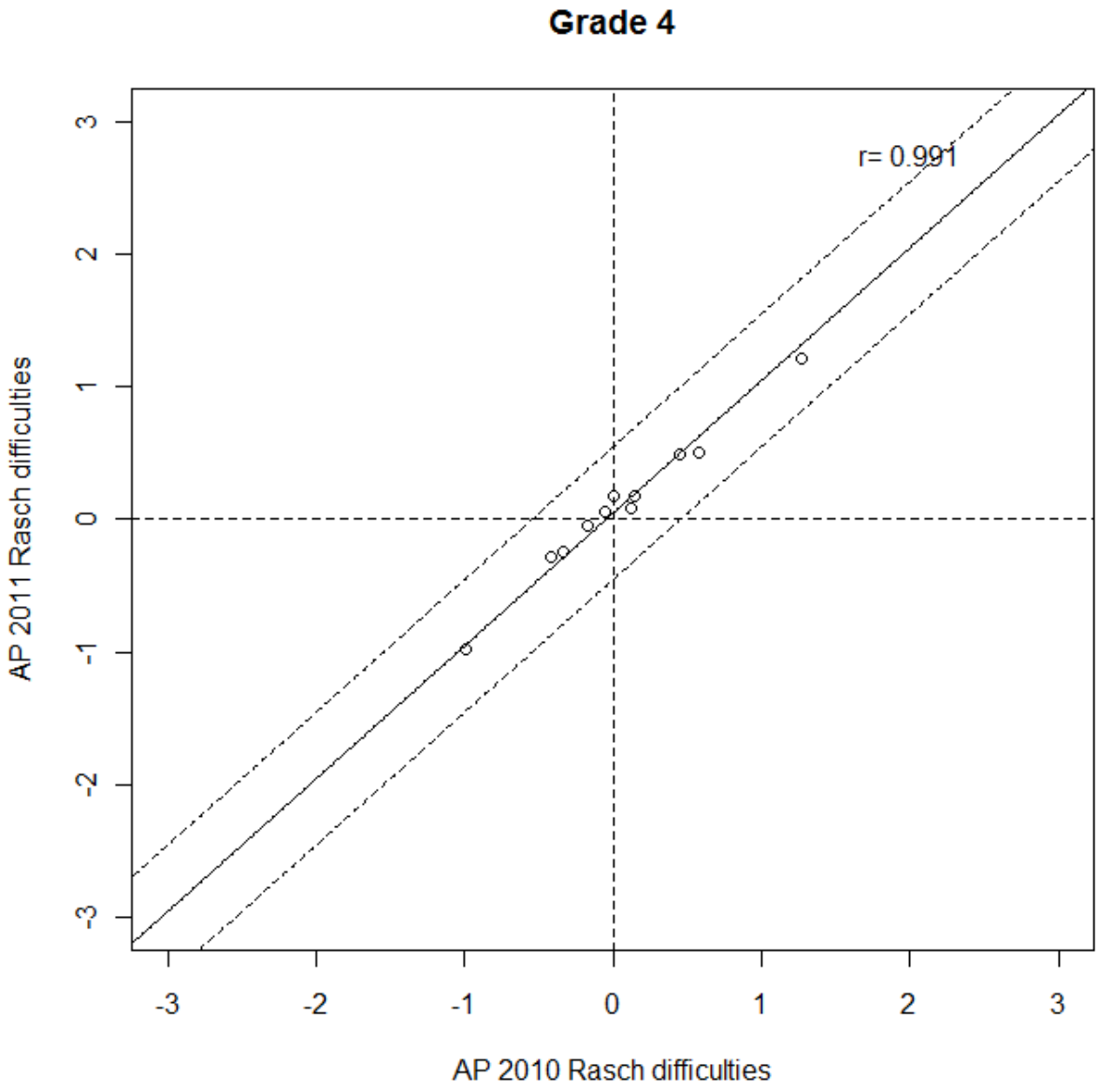
Wyse, A. E., & Hao, S. (2012). An evaluation of item response theory classification accuracy and consistency indices. *Applied Psychological Measurement*, *36(7),* 602-624. doi: 10.1177/0013164410395934.

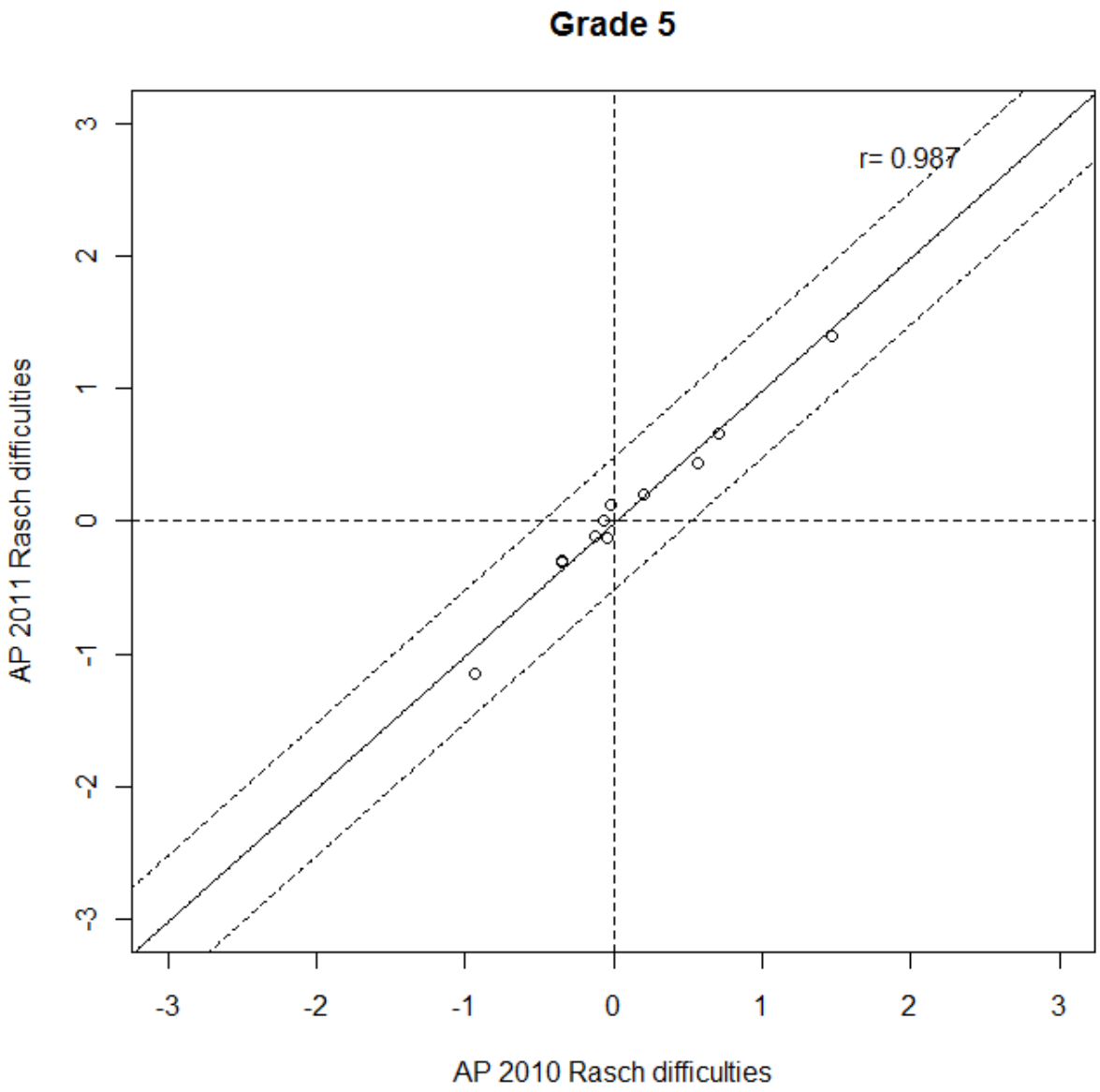
Wyse, A. E., Zeng, J., & Martineau, J. A. (2011). A graphical transition table for communicating status and growth. *Practical Assessment, Research and Evaluation, 16(11).* Retrieved from http://pareonline.net/getvn.asp?v=16&n=11.

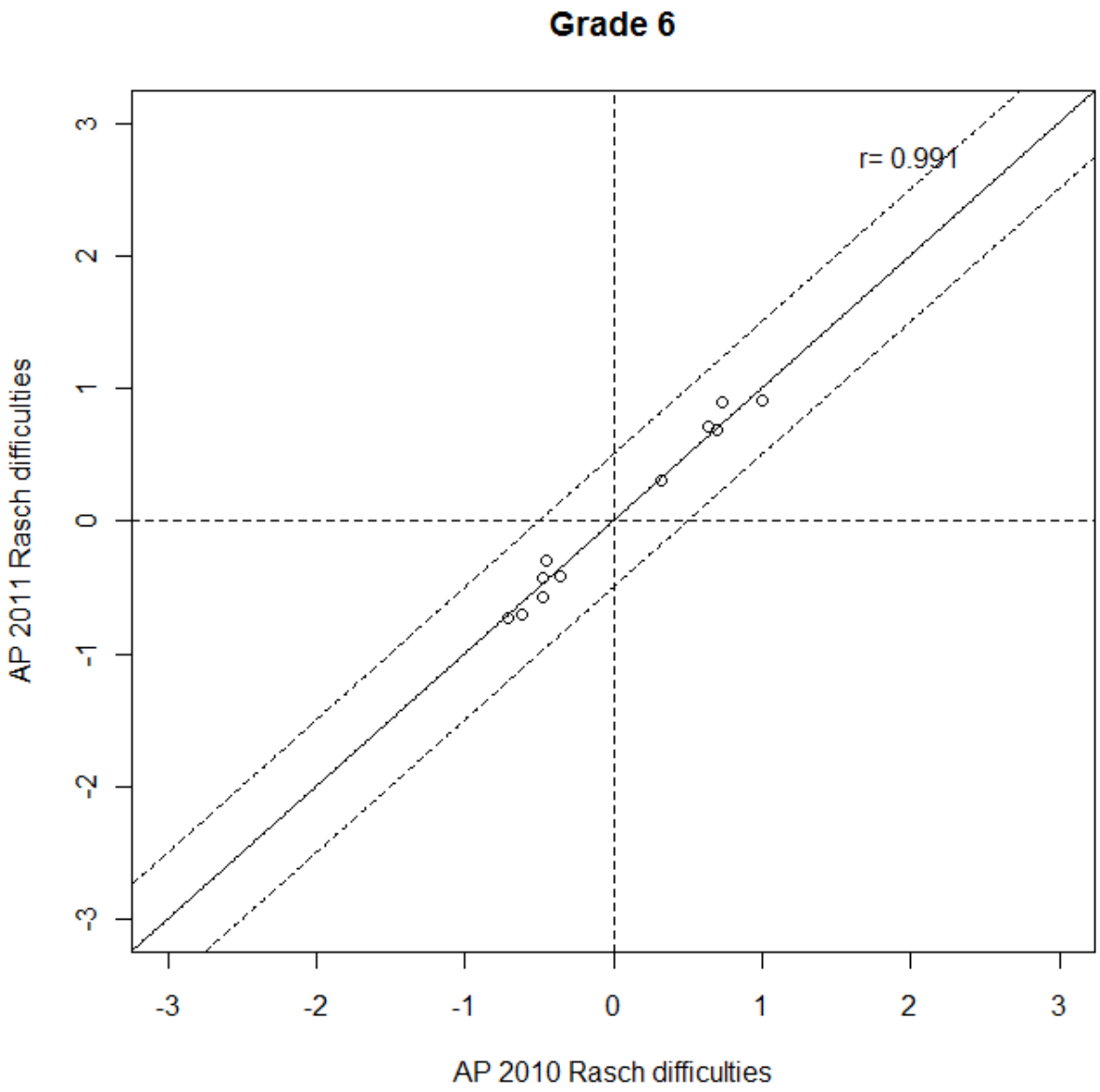
Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–347). Hillsdale, NJ: Erlbaum.

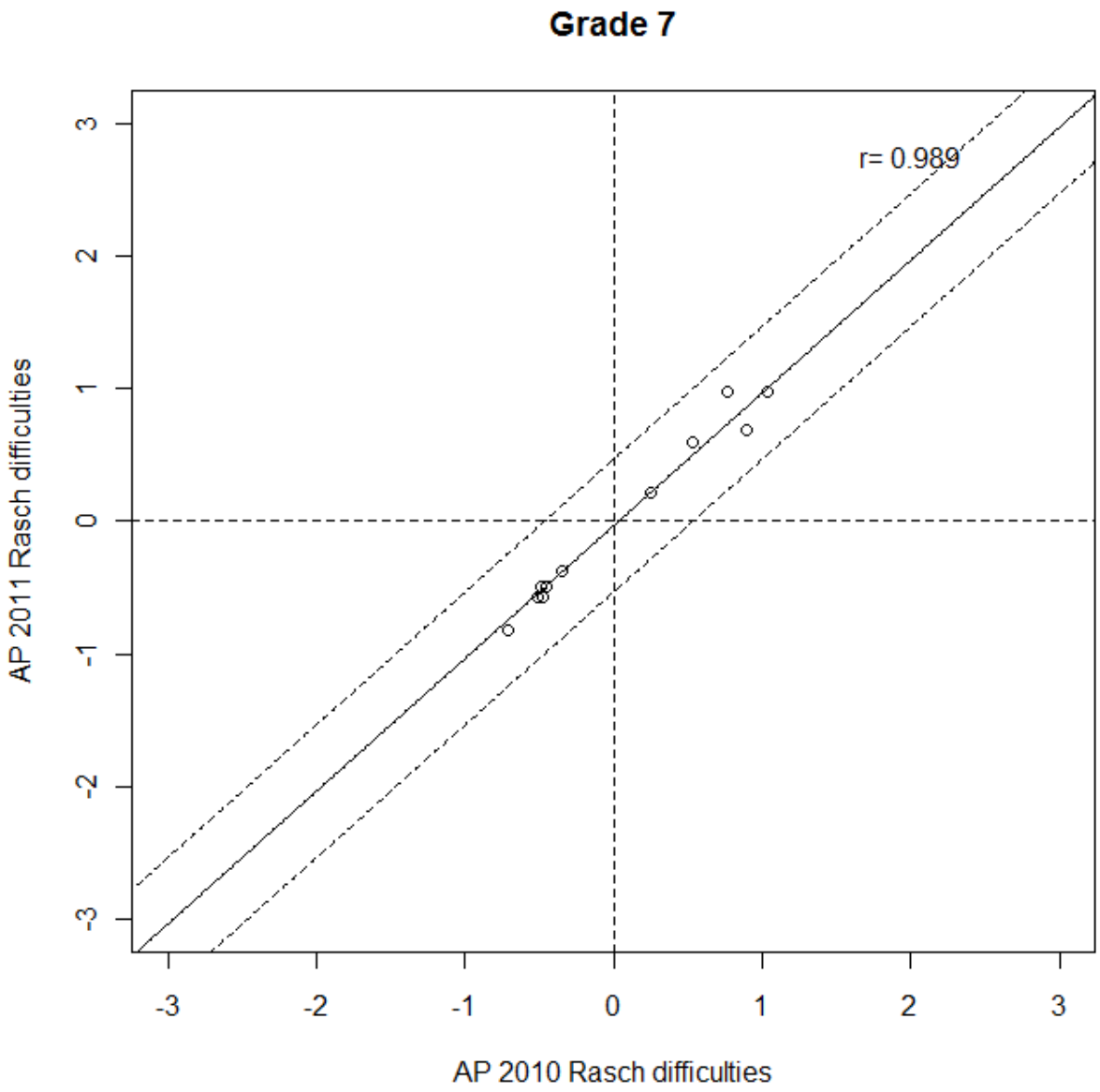
APPENDIX A: ANCHOR ITEM PLOTS

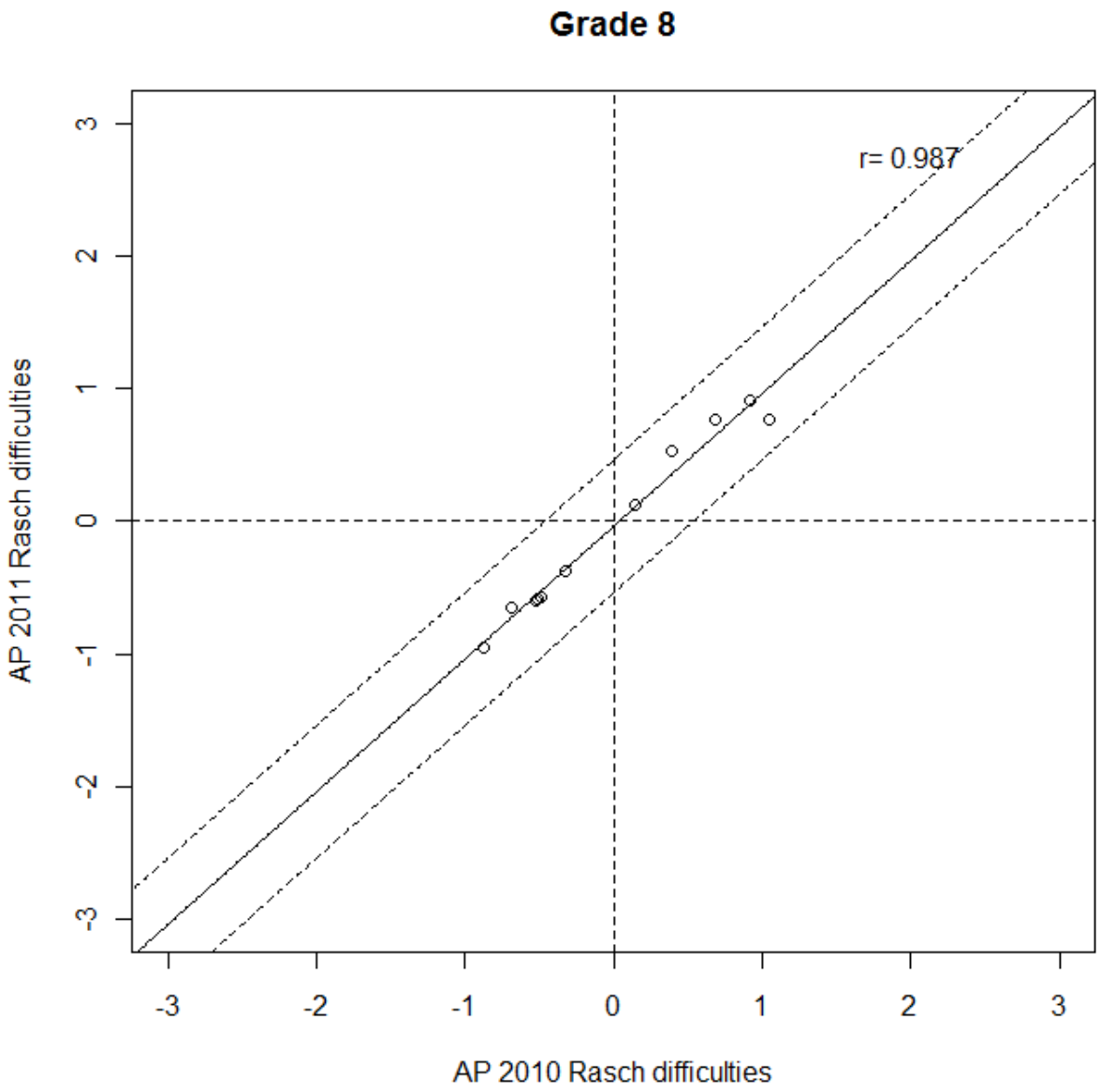

Grade 3
r= 0.964
AP 2011 Rasch difficulties
AP 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

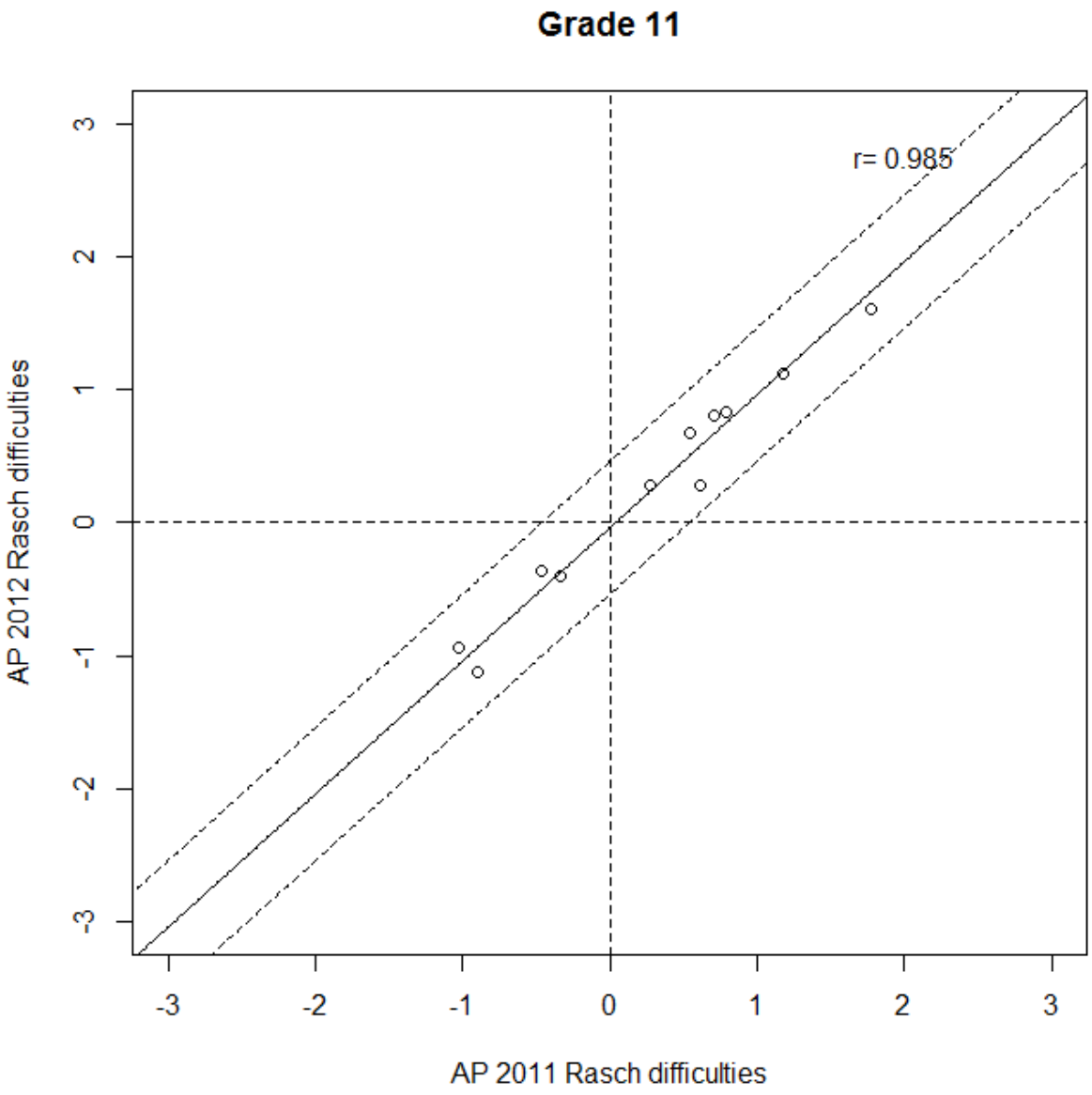

Grade 4
r= 0.991
AP 2011 Rasch difficulties
AP 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

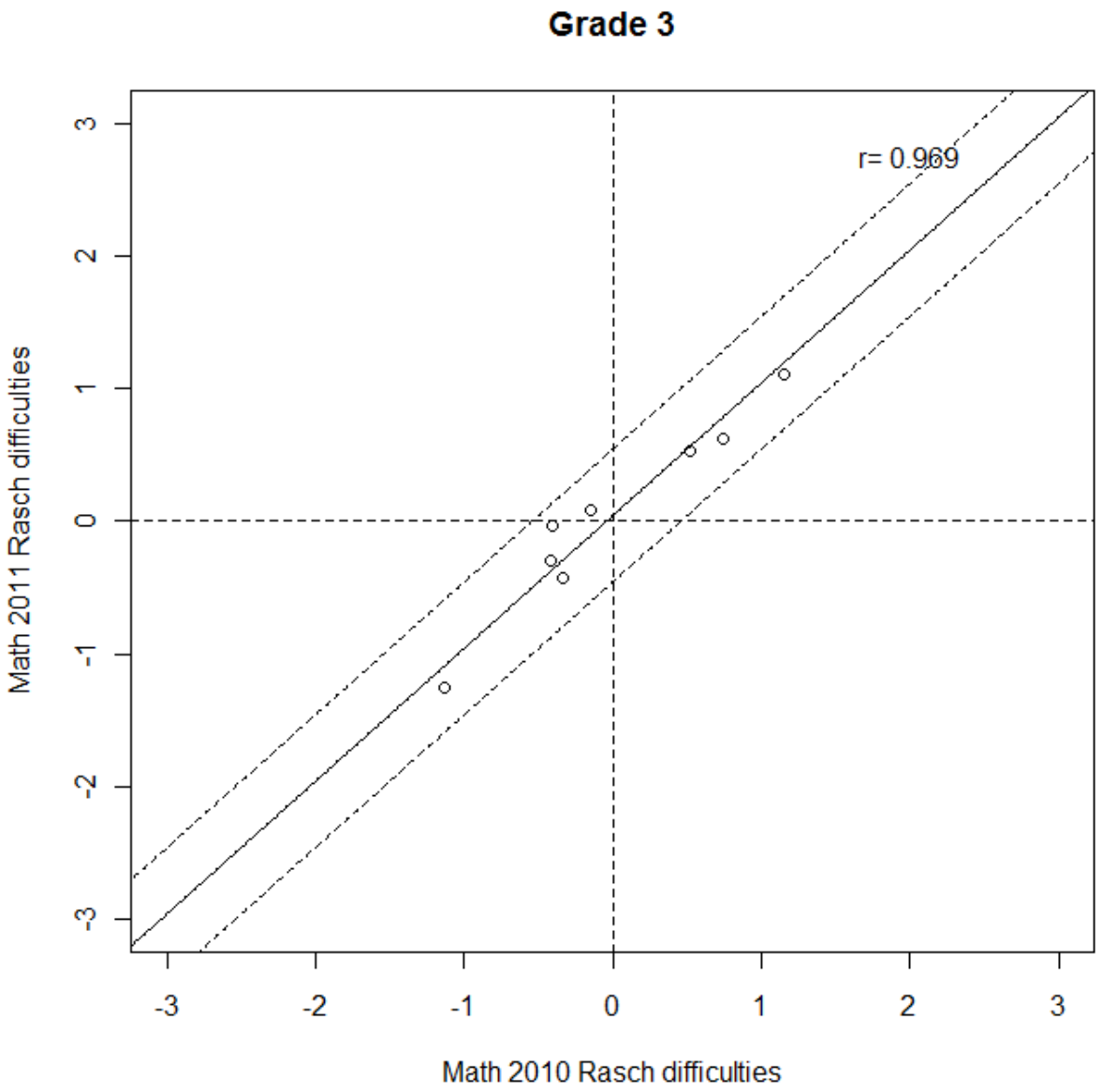

Grade 5
r= 0.987
AP 2011 Rasch difficulties
AP 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

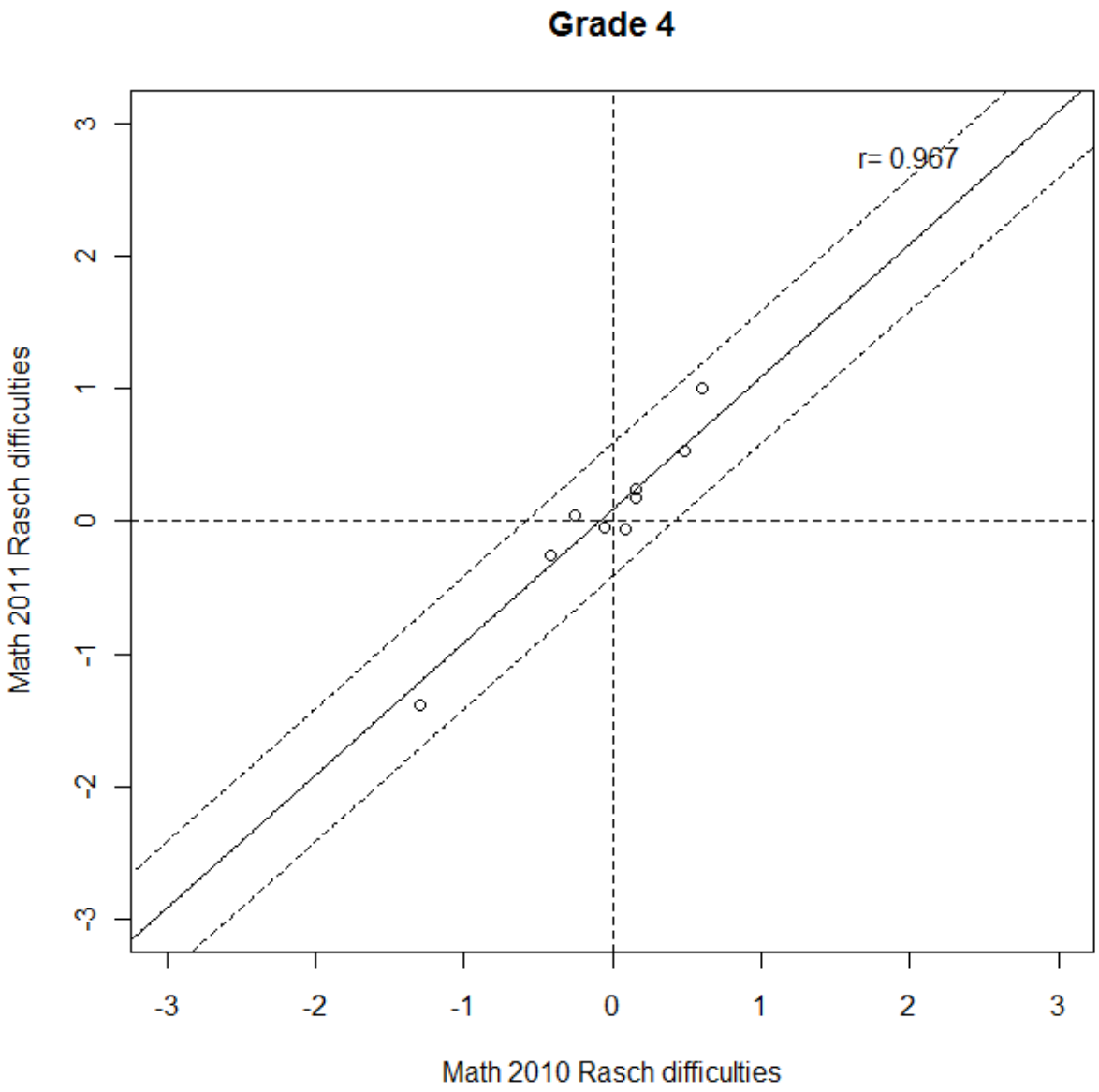

Grade 6
r= 0.991
AP 2011 Rasch difficulties
AP 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

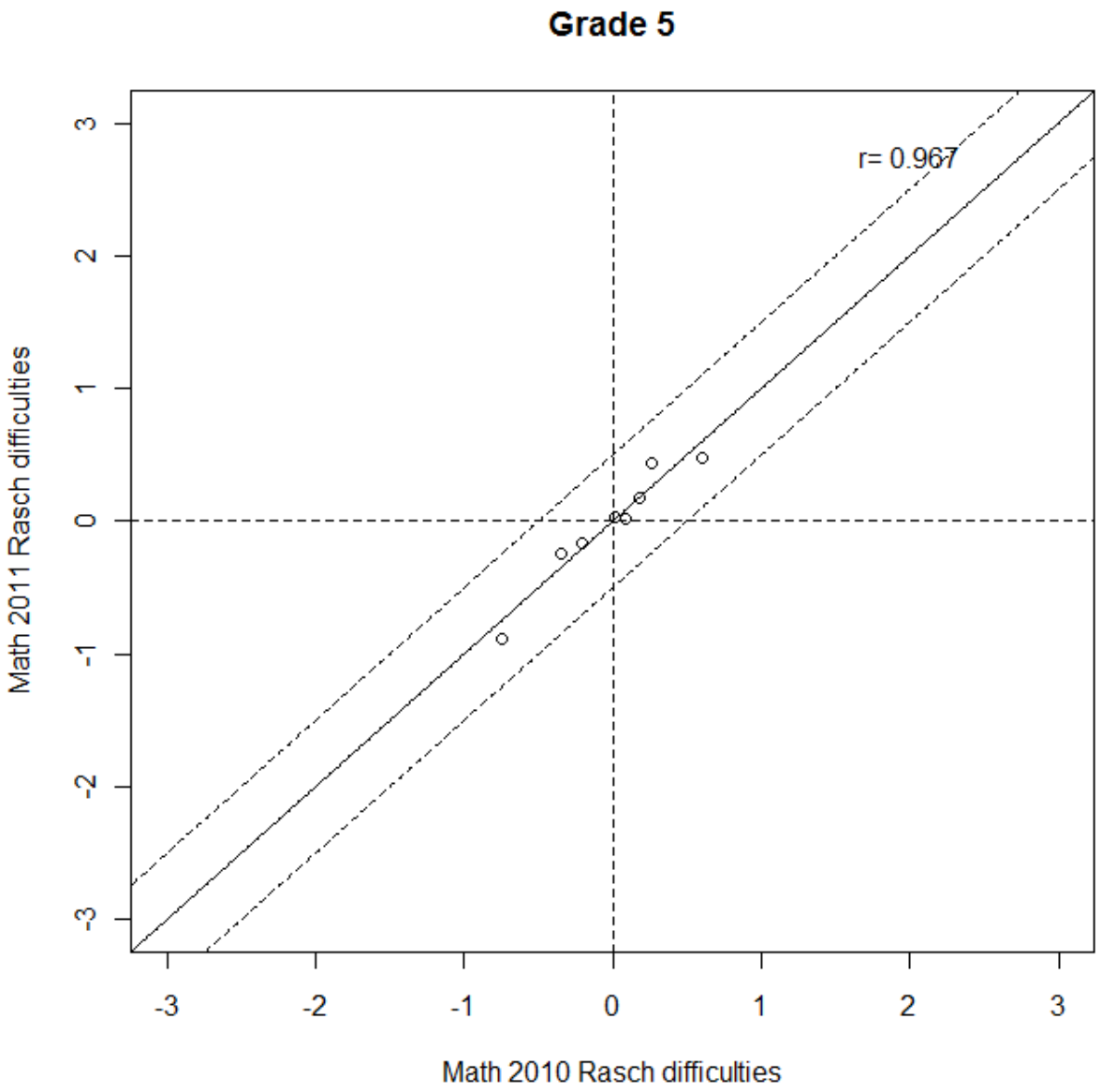

Grade 7
r= 0.989
AP 2011 Rasch difficulties
AP 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

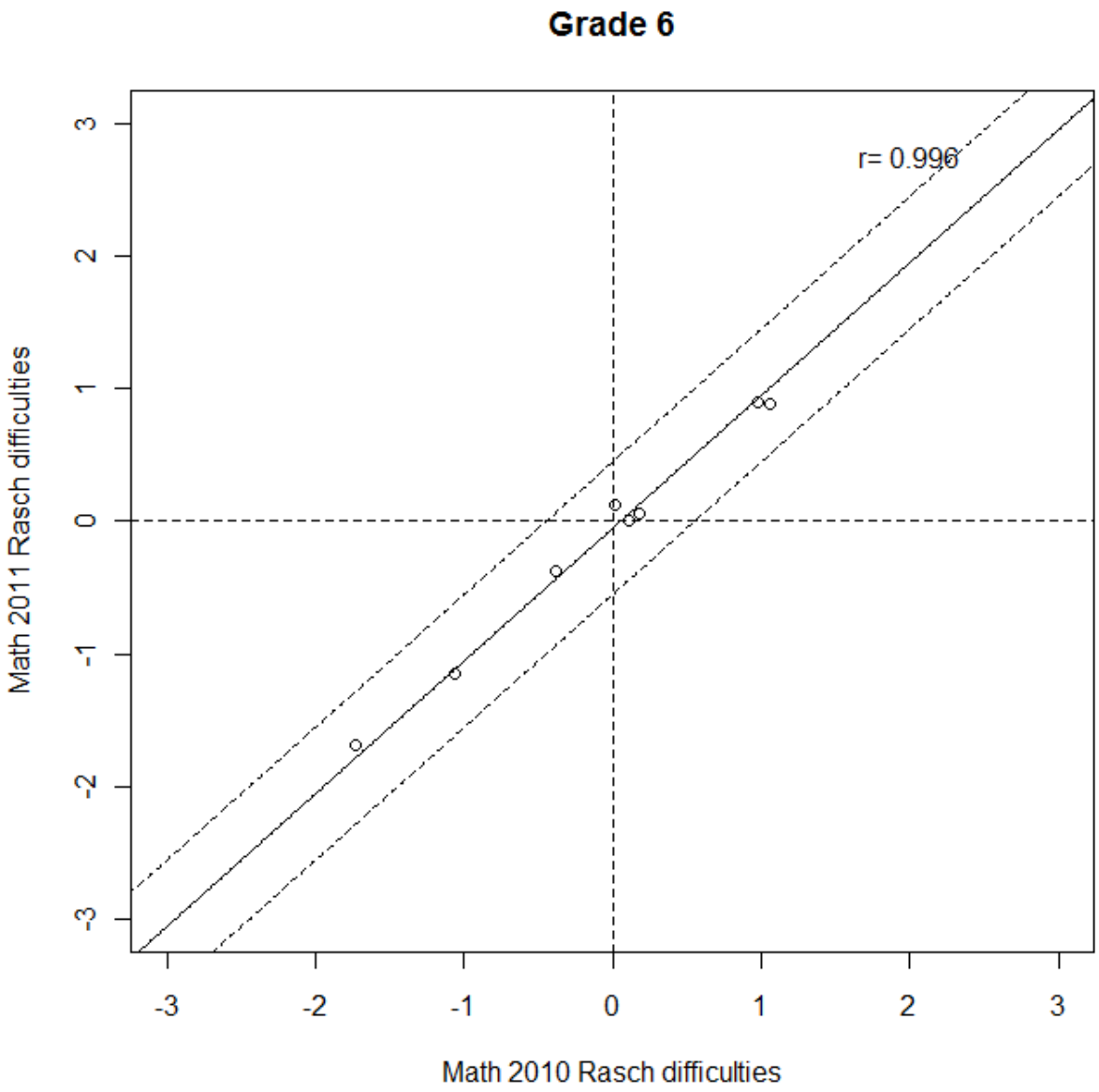

Grade 8
r= 0.987
AP 2011 Rasch difficulties
AP 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

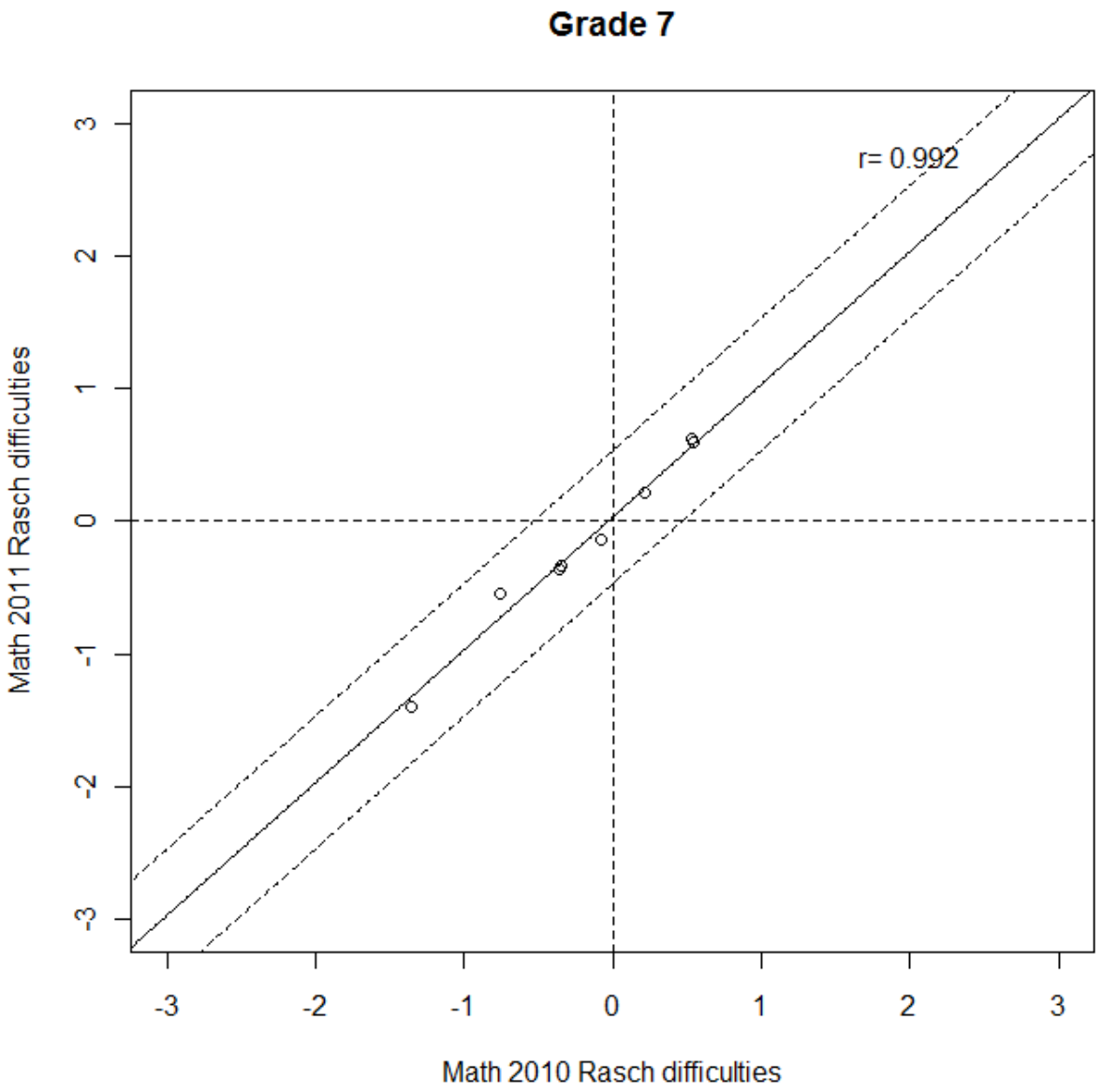

Grade 11
r= 0.985
AP 2012 Rasch difficulties
AP 2011 Rasch difficulties
-3
-2
-1
0
1
2
3

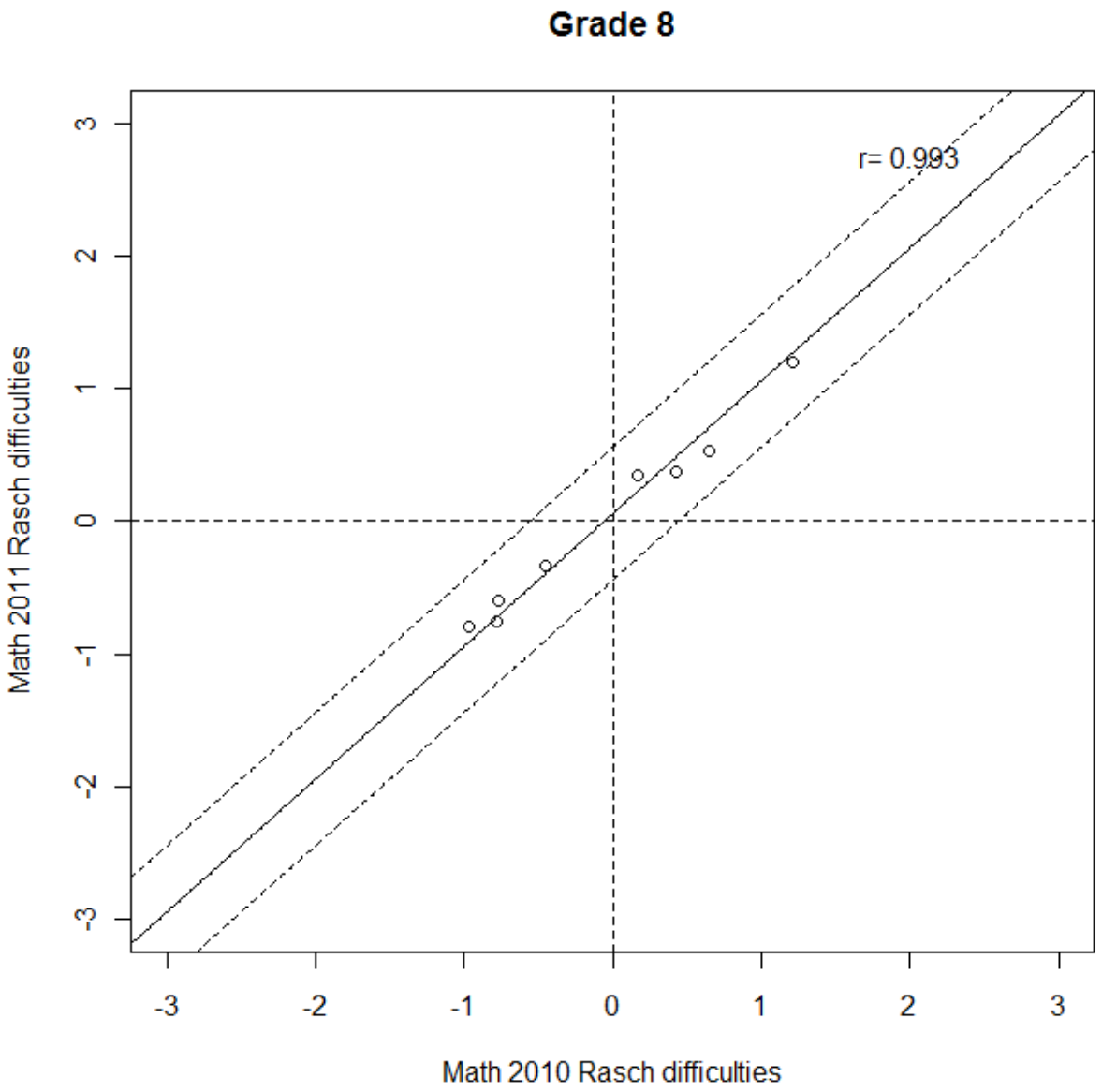

Grade 3
r= 0.969
Math 2011 Rasch difficulties
Math 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

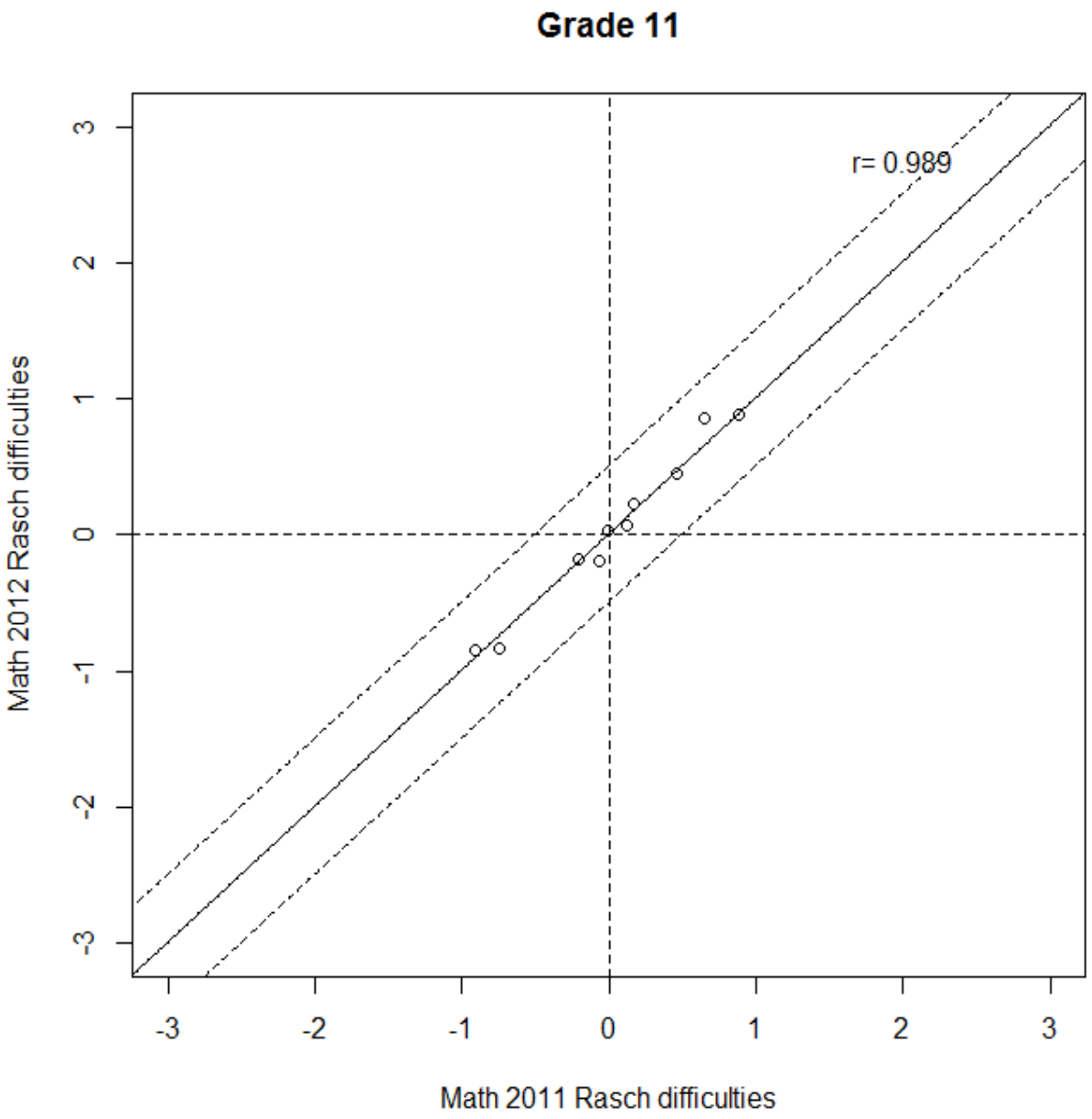

Grade 4
r= 0.967
Math 2011 Rasch difficulties
Math 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

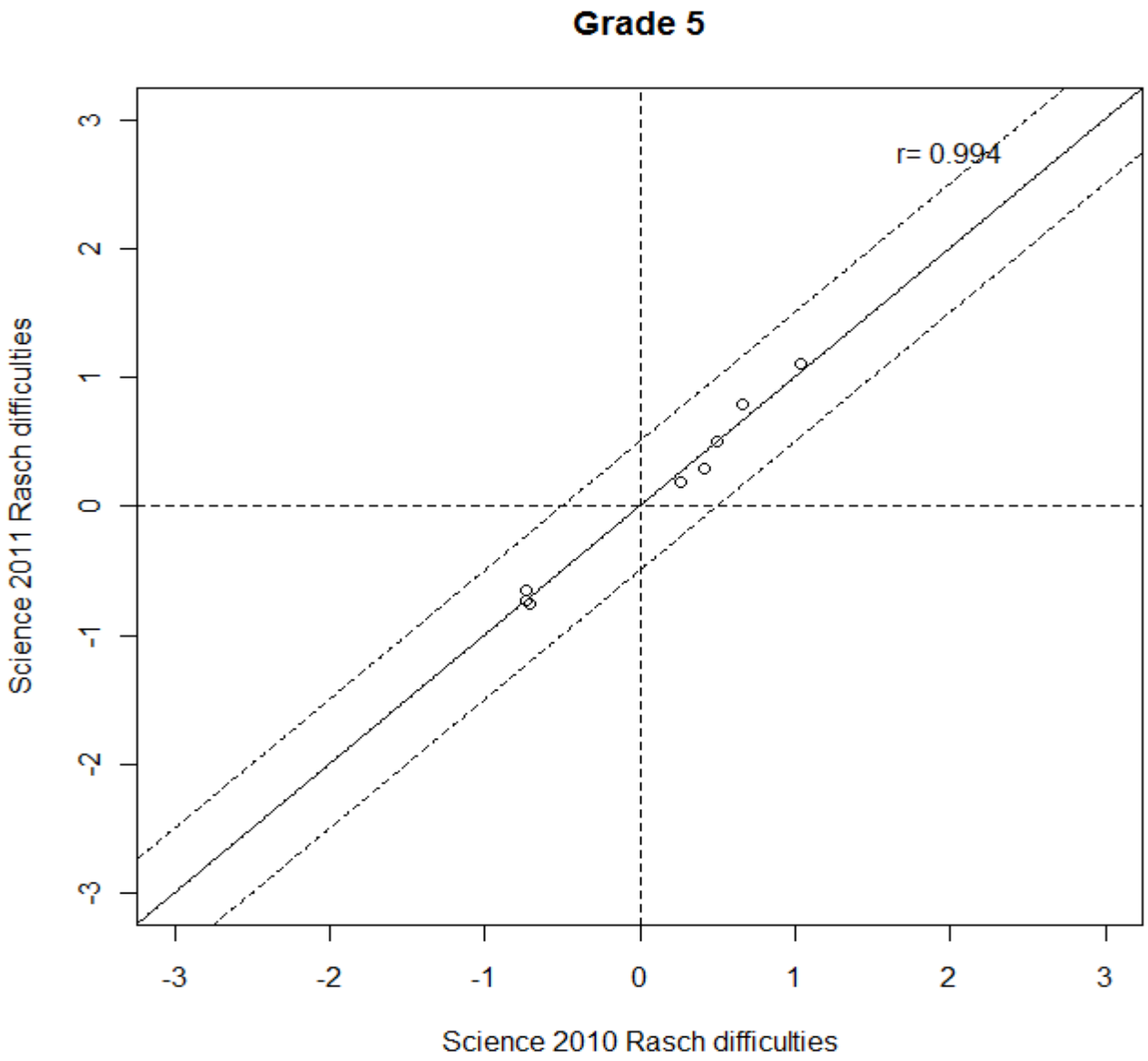

Grade 5
r= 0.967
Math 2011 Rasch difficulties
Math 2010 Rasch difficulties
-3
-2
-1
0
1
2
3


Grade 6
r= 0.996
Math 2011 Rasch difficulties
Math 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

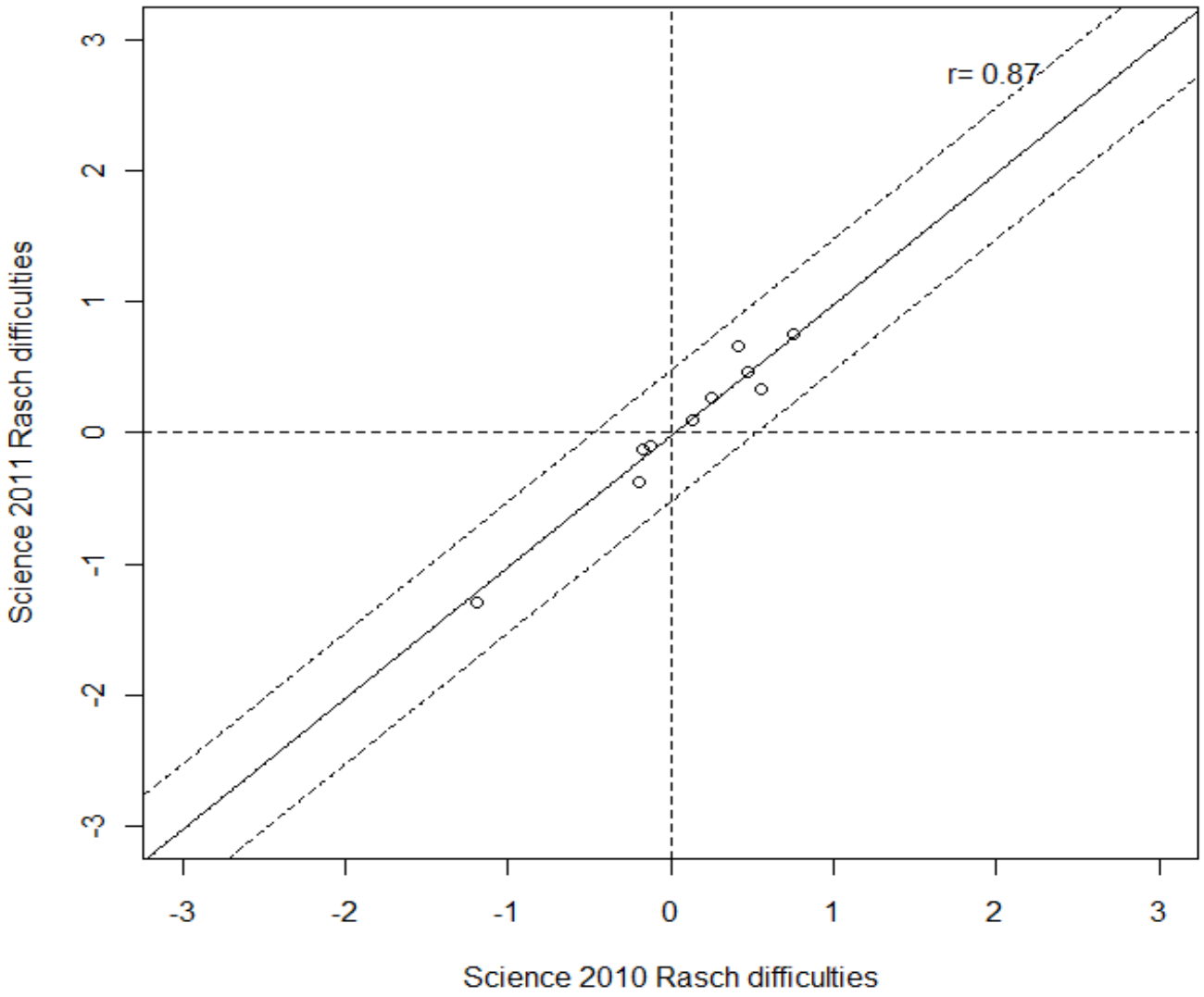

Grade 7
r= 0.992
Math 2011 Rasch difficulties
Math 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

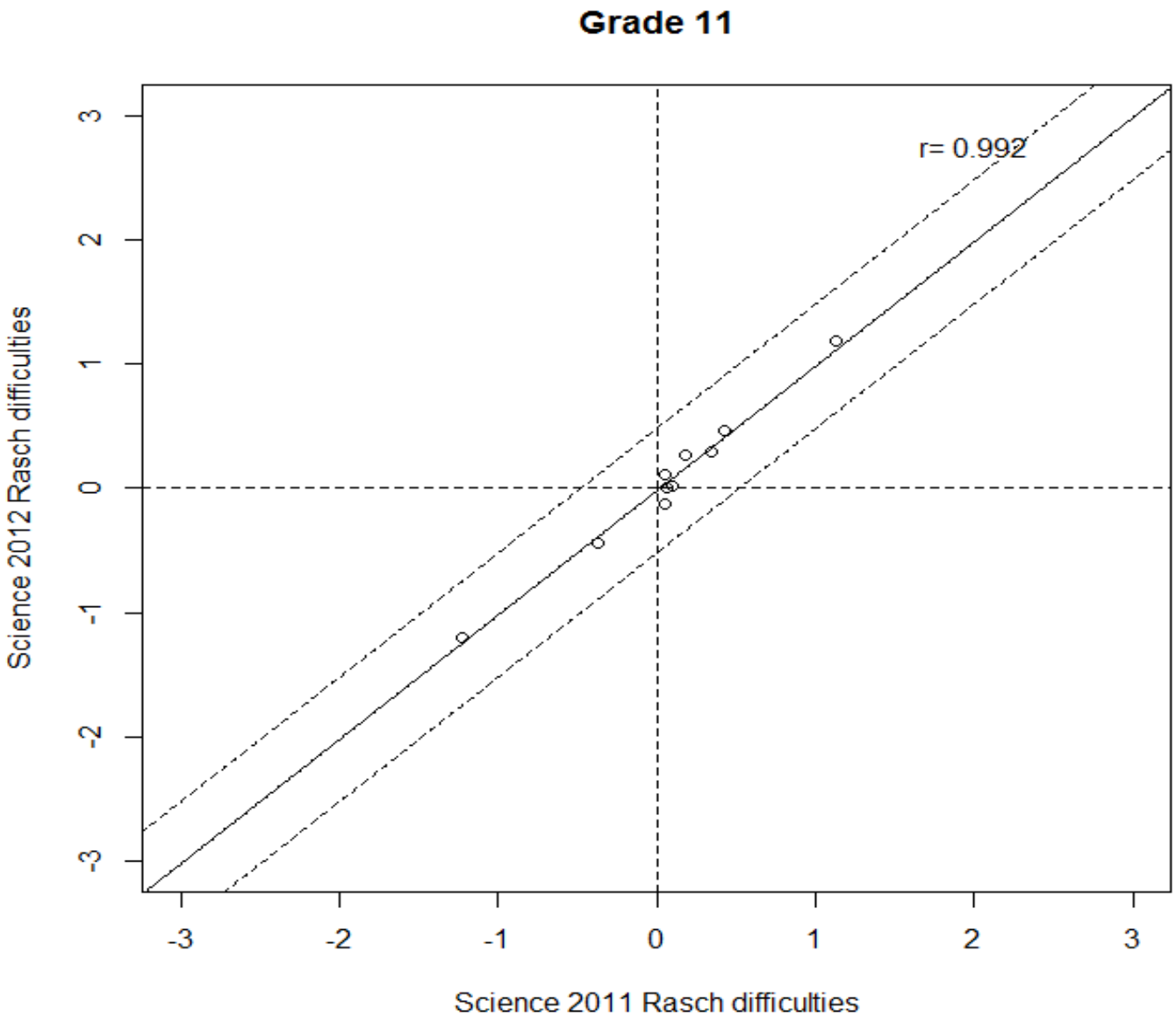

Grade 8
r= 0.993
Math 2011 Rasch difficulties
Math 2010 Rasch difficulties
-3
-2
-1
0
1
2
3


Grade 11
r= 0.989
Math 2012 Rasch difficulties
Math 2011 Rasch difficulties
-3
-2
-1
0
1
2
3

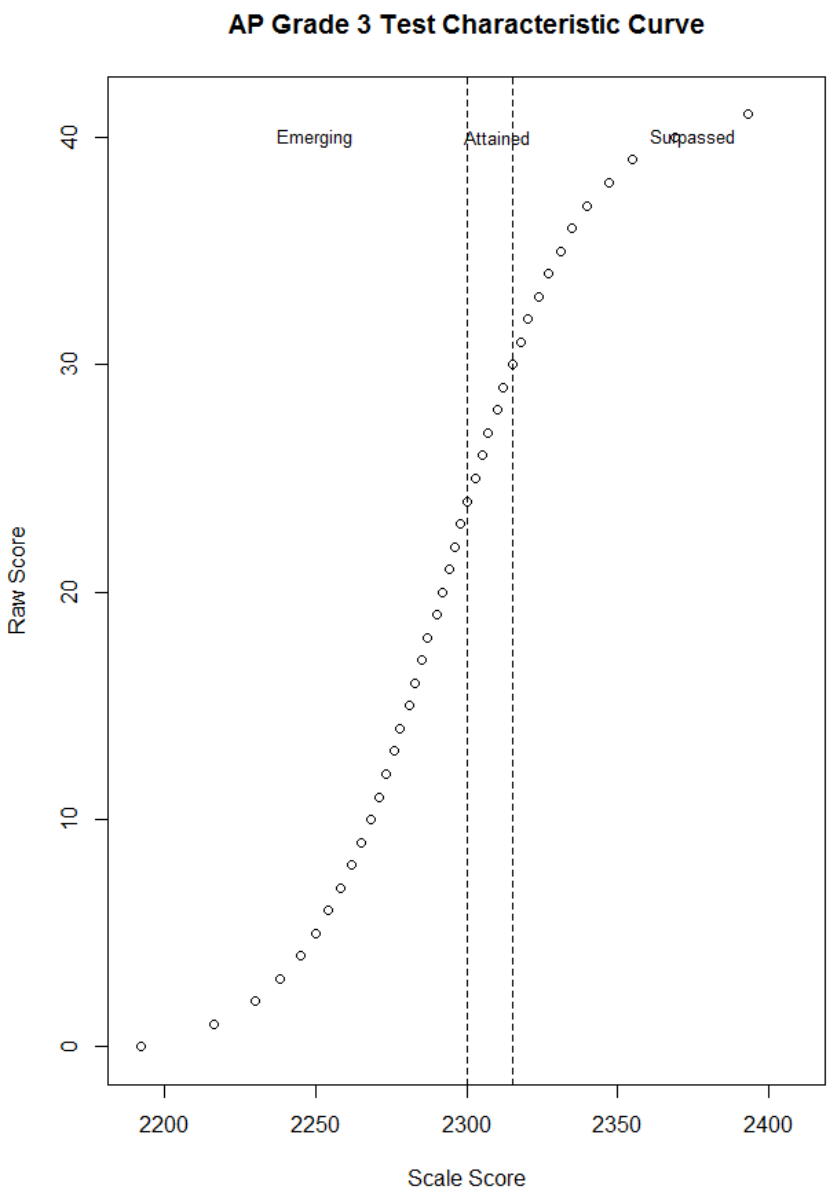

Grade 5
r= 0.994
Science 2011 Rasch difficulties
Science 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

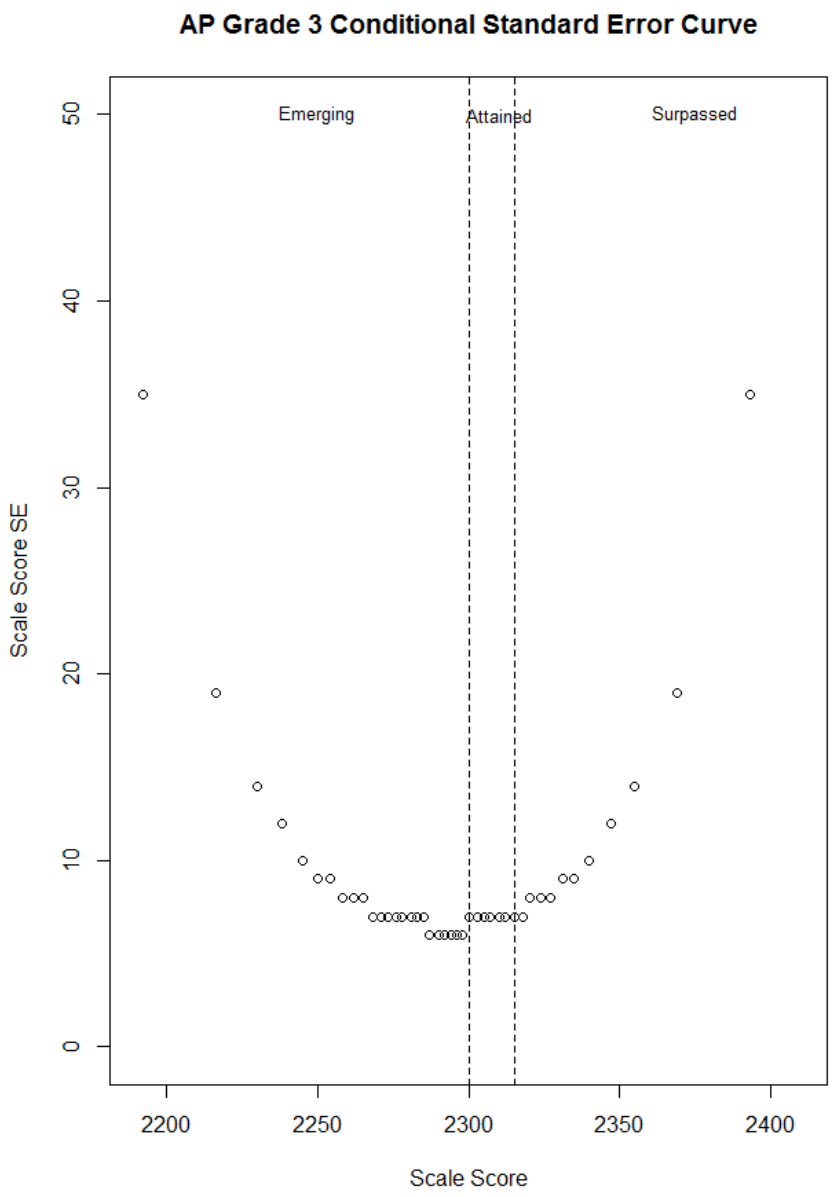
Grade 8

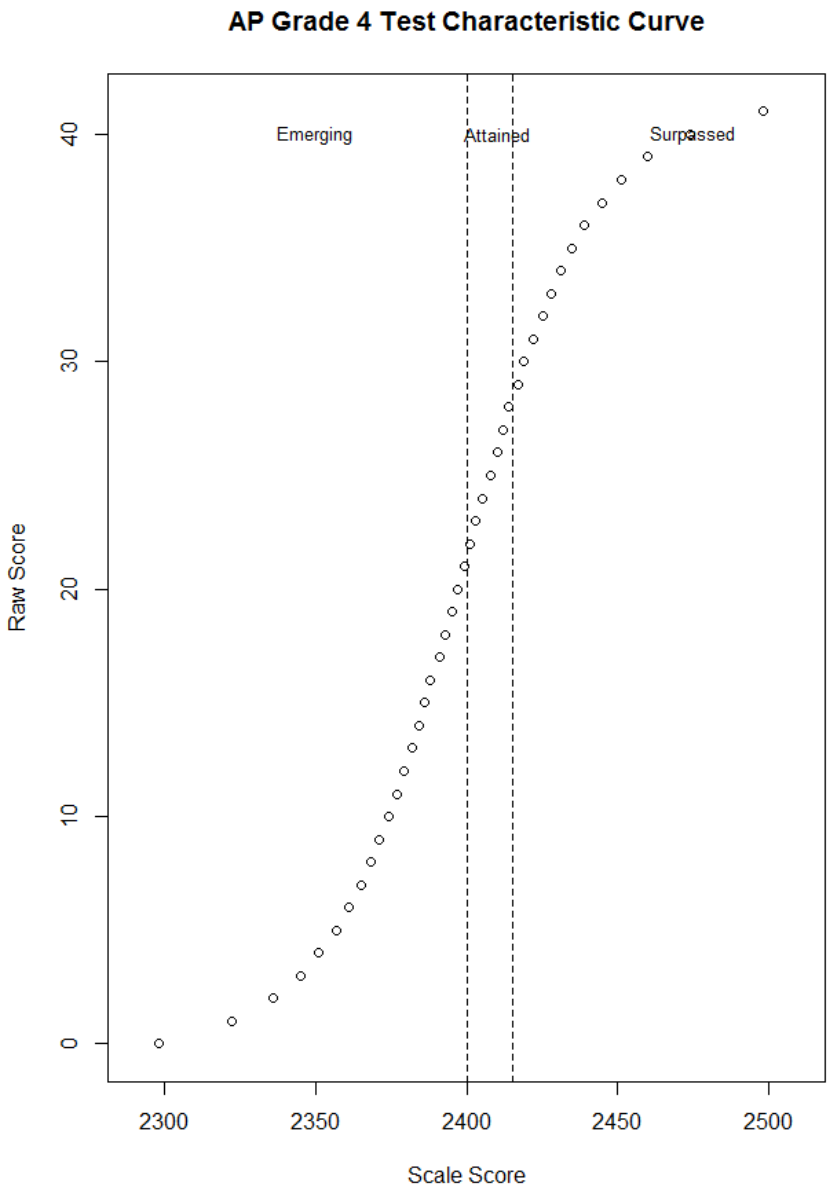

r= 0.87
Science 2011 Rasch difficulties
Science 2010 Rasch difficulties
-3
-2
-1
0
1
2
3

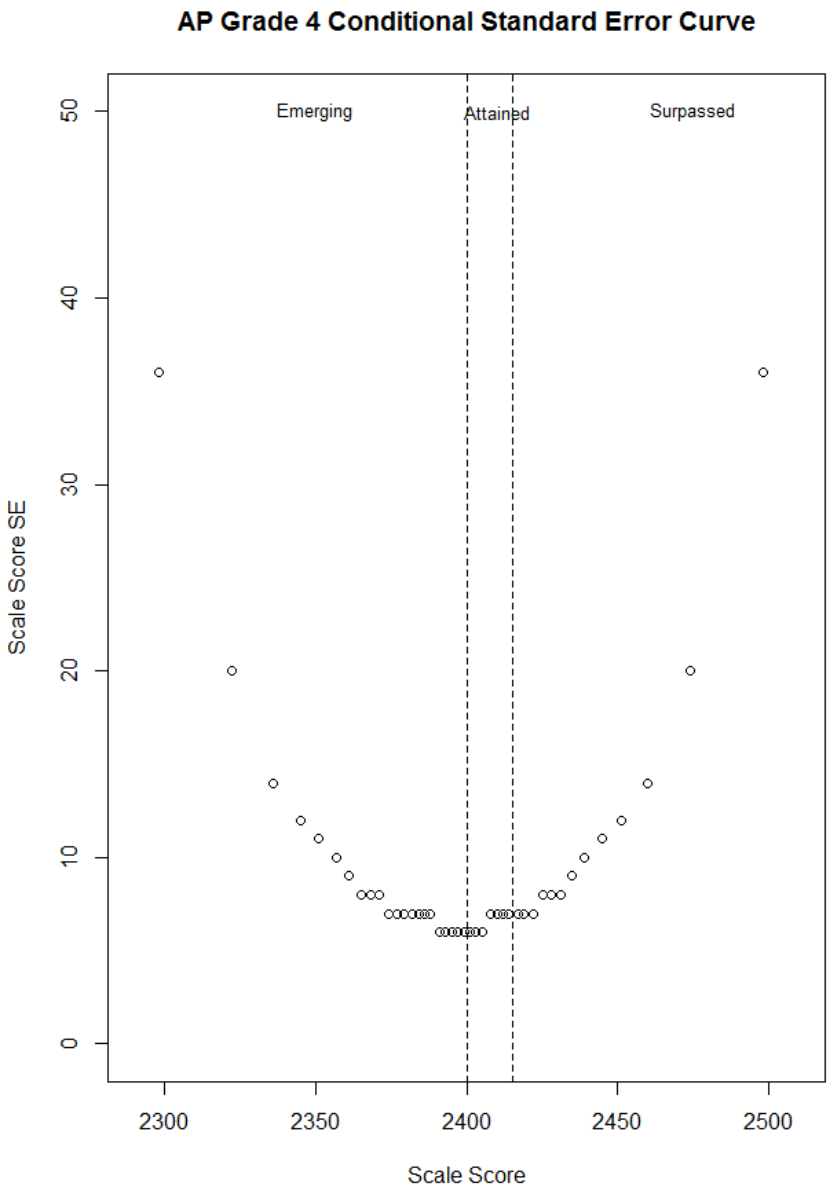

Grade 11
r= 0.992
Science 2012 Rasch difficulties
Science 2011 Rasch difficulties
-3
-2
-1
0
1
2
3

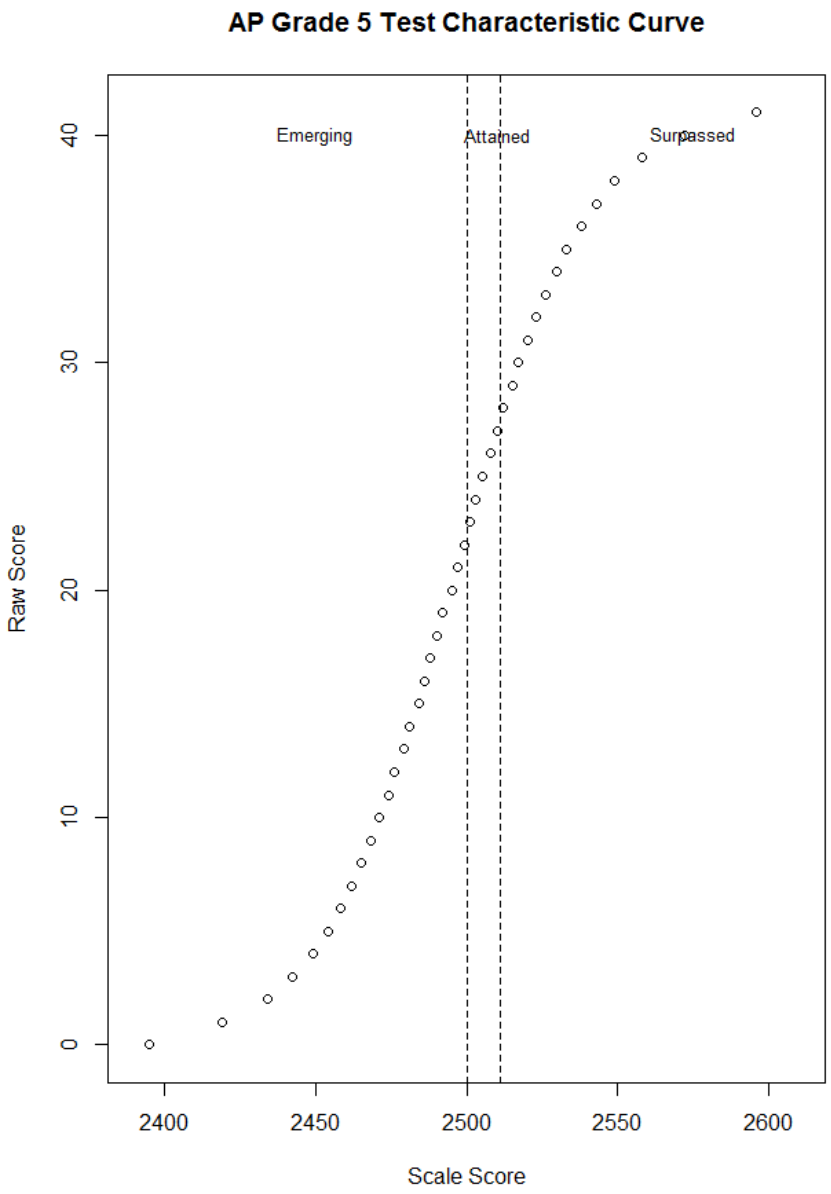
APPENDIX B:
TEST CHARACTERISTIC CURVES AND STANDARD ERROR CURVES

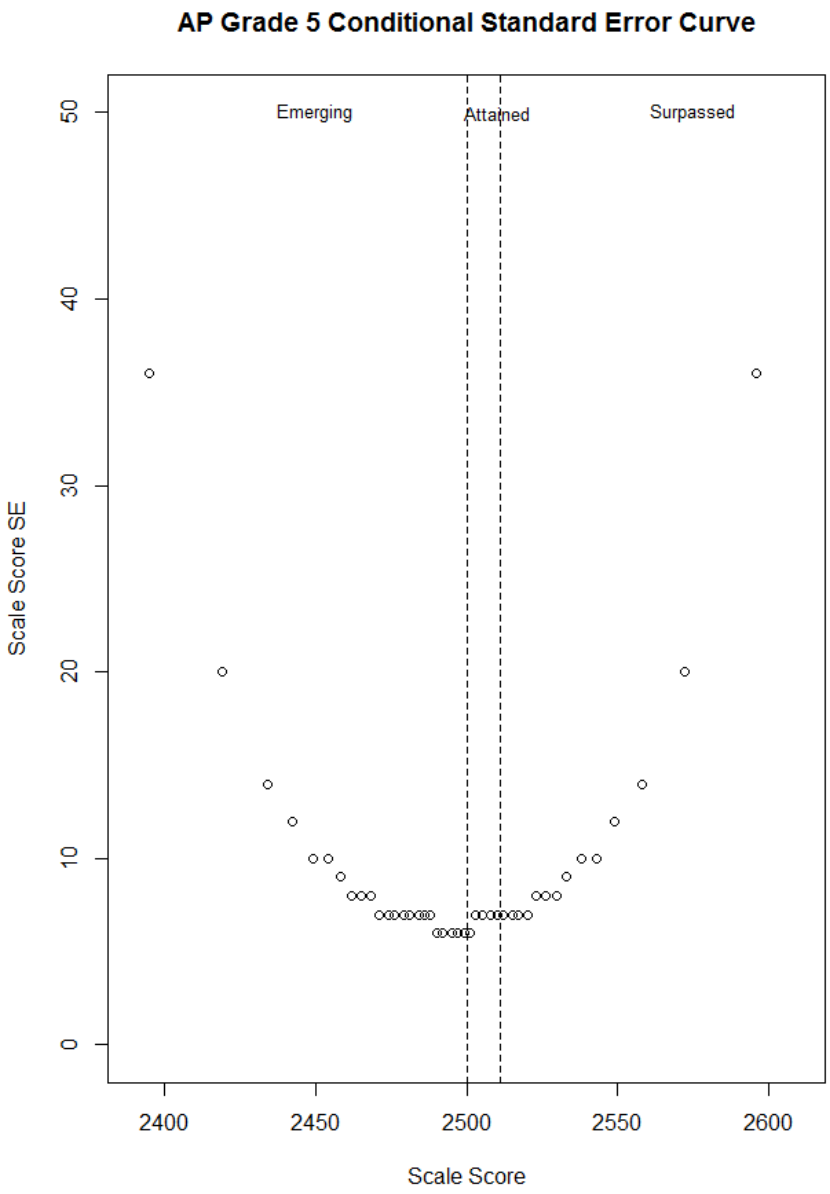

AP Grade 3 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
Scale Score
0
10
20
30
40
2200
2250
2300
2350
2400

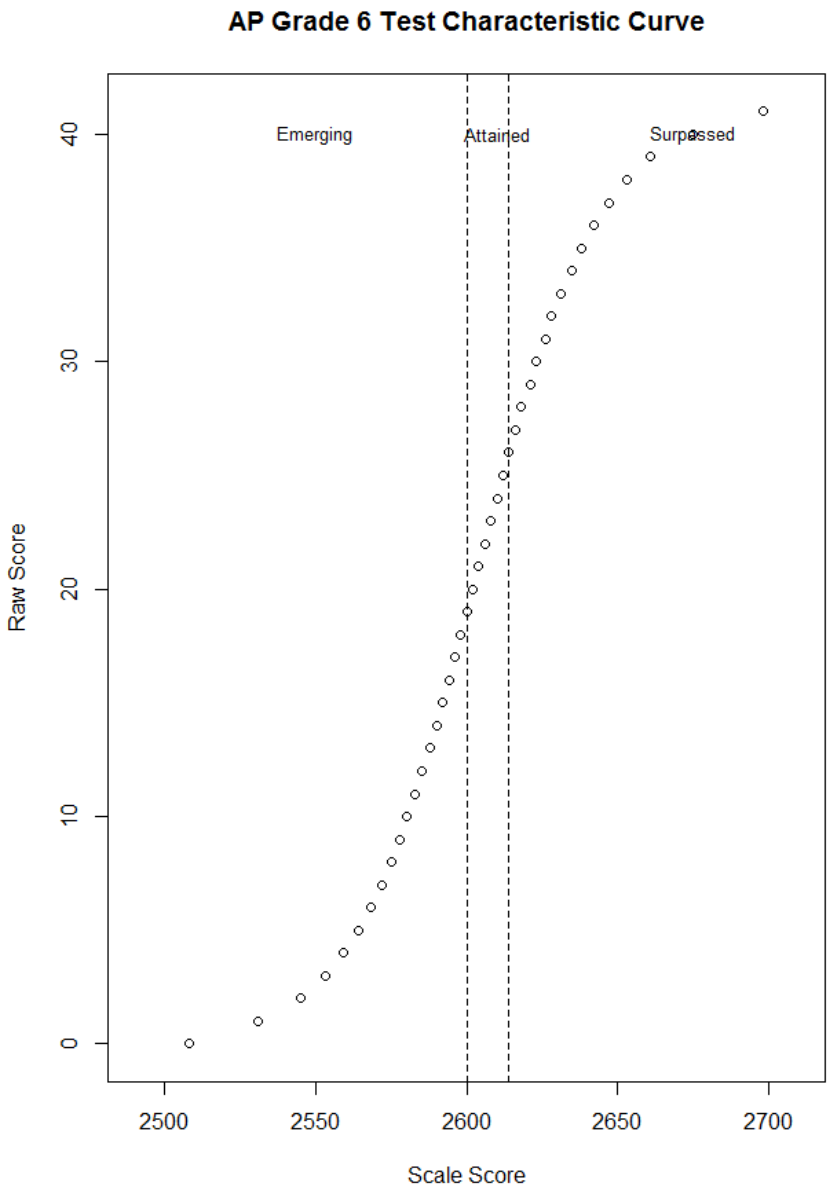

AP Grade 3 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
Scale Score
0
10
20
30
40
50
2200
2250
2300
2350
2400

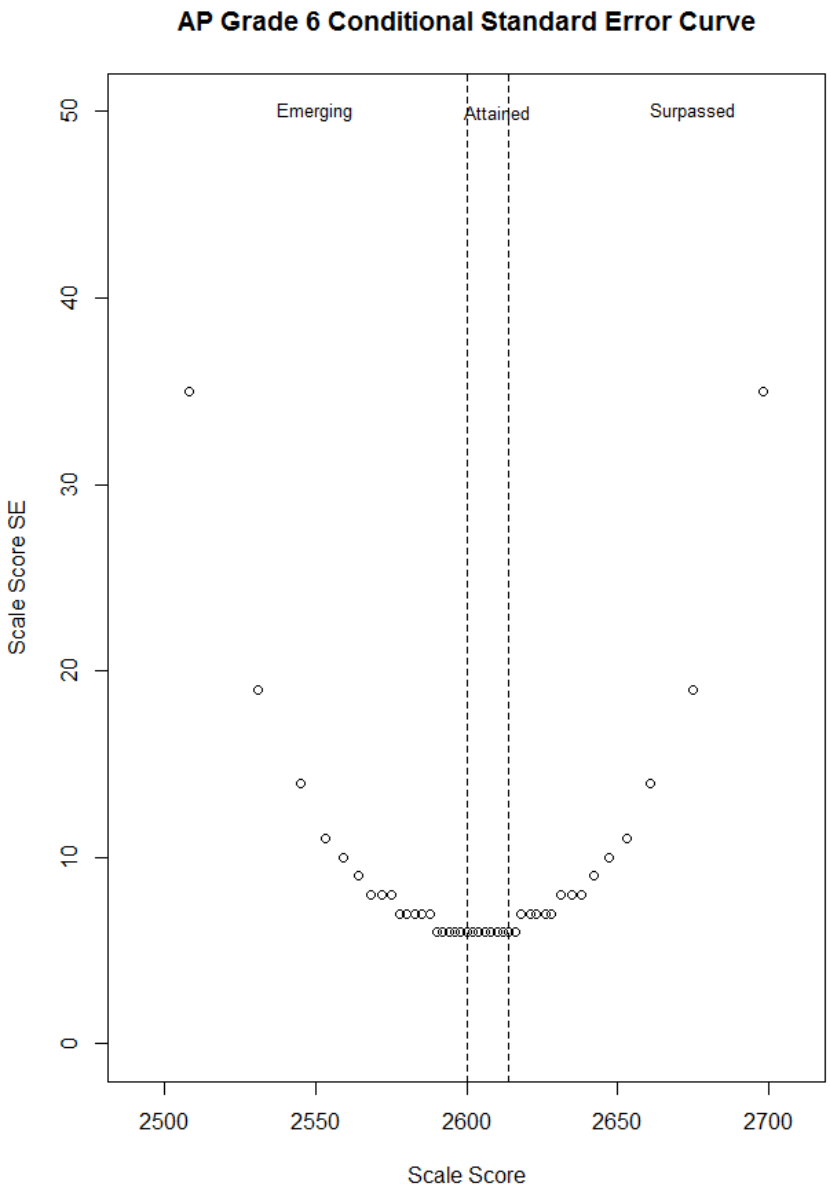

AP Grade 4 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
10
20
30
40
2300
2350
2400
2450
2500
Scale Score

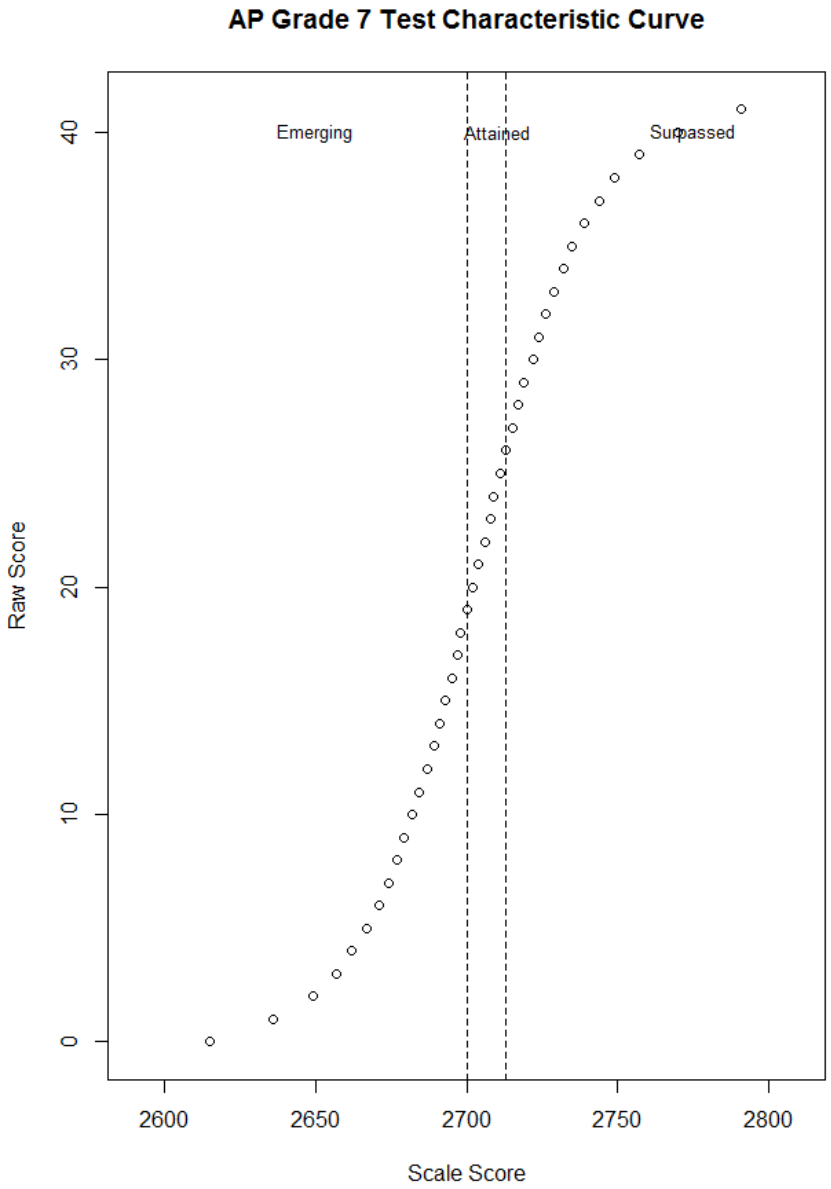

AP Grade 4 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
2300
2350
2400
2450
2500
Scale Score

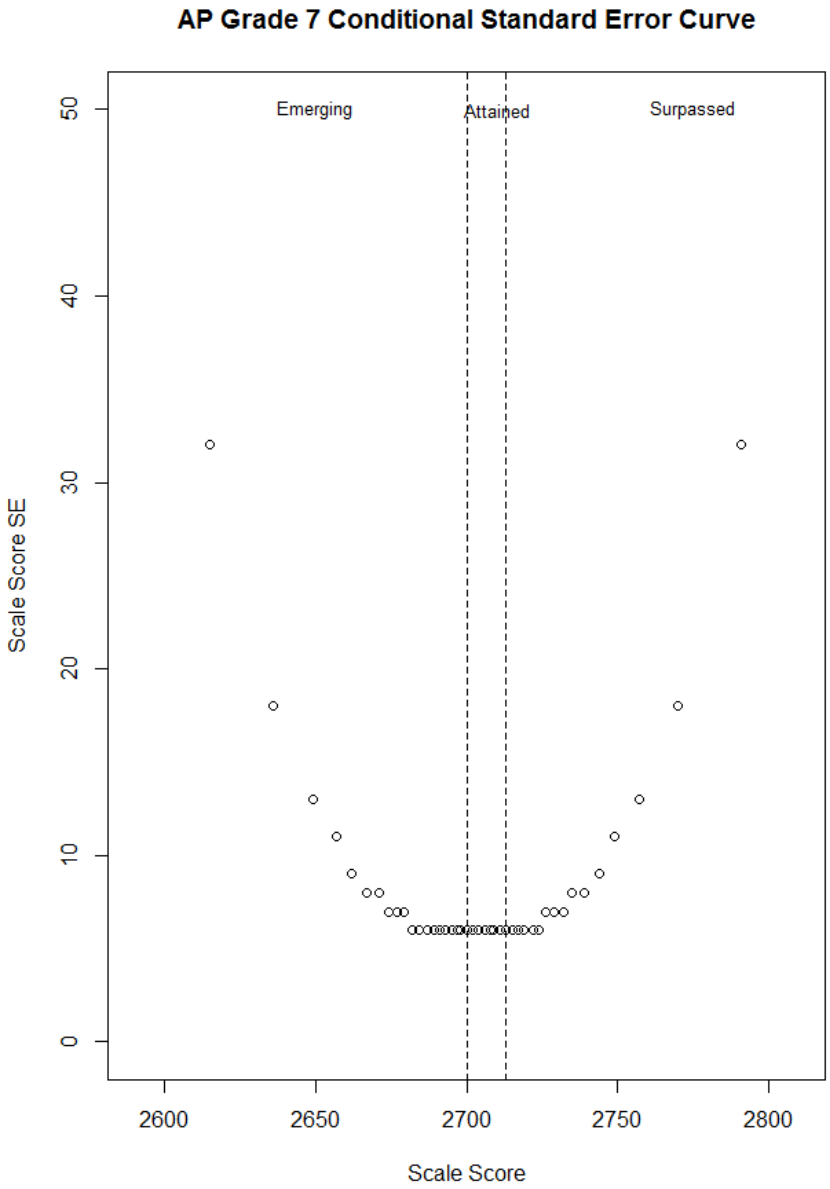

AP Grade 5 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
10
20
30
40
2400
2450
2500
2550
2600
Scale Score

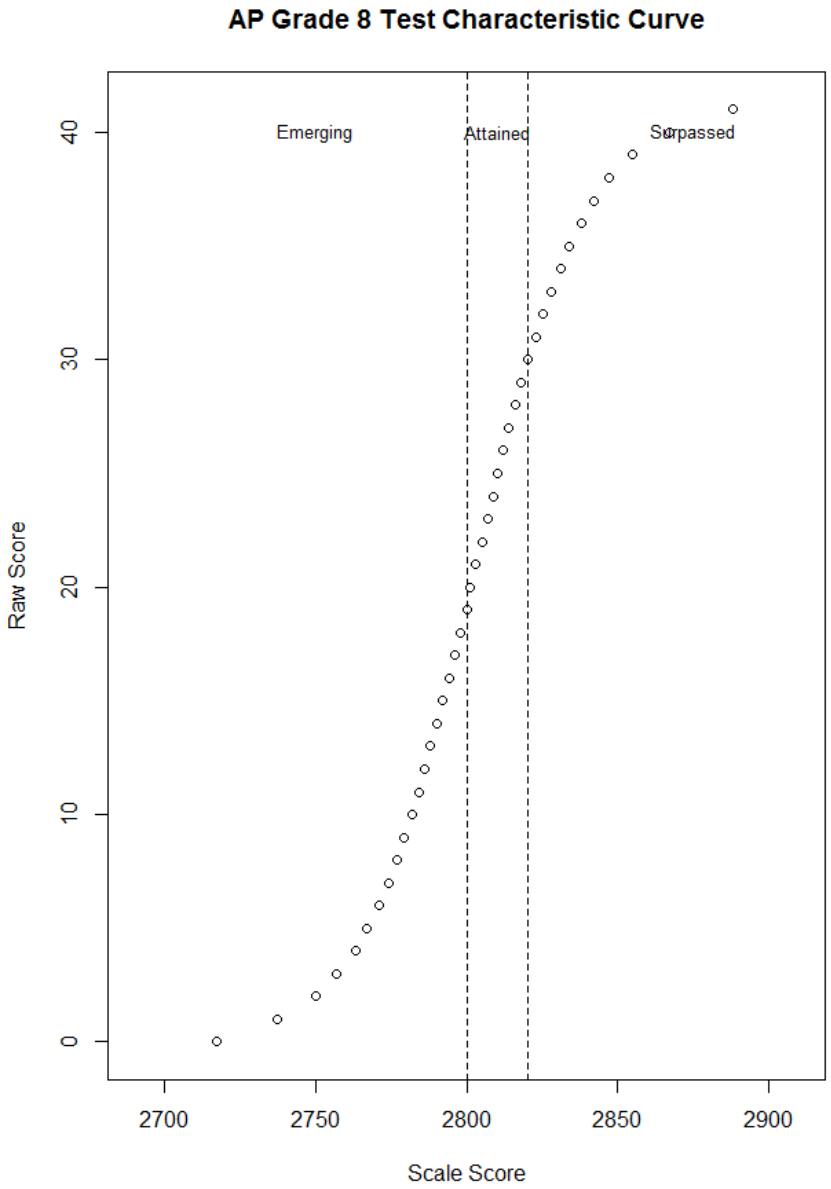

AP Grade 5 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
2400
2450
2500
2550
2600
Scale Score

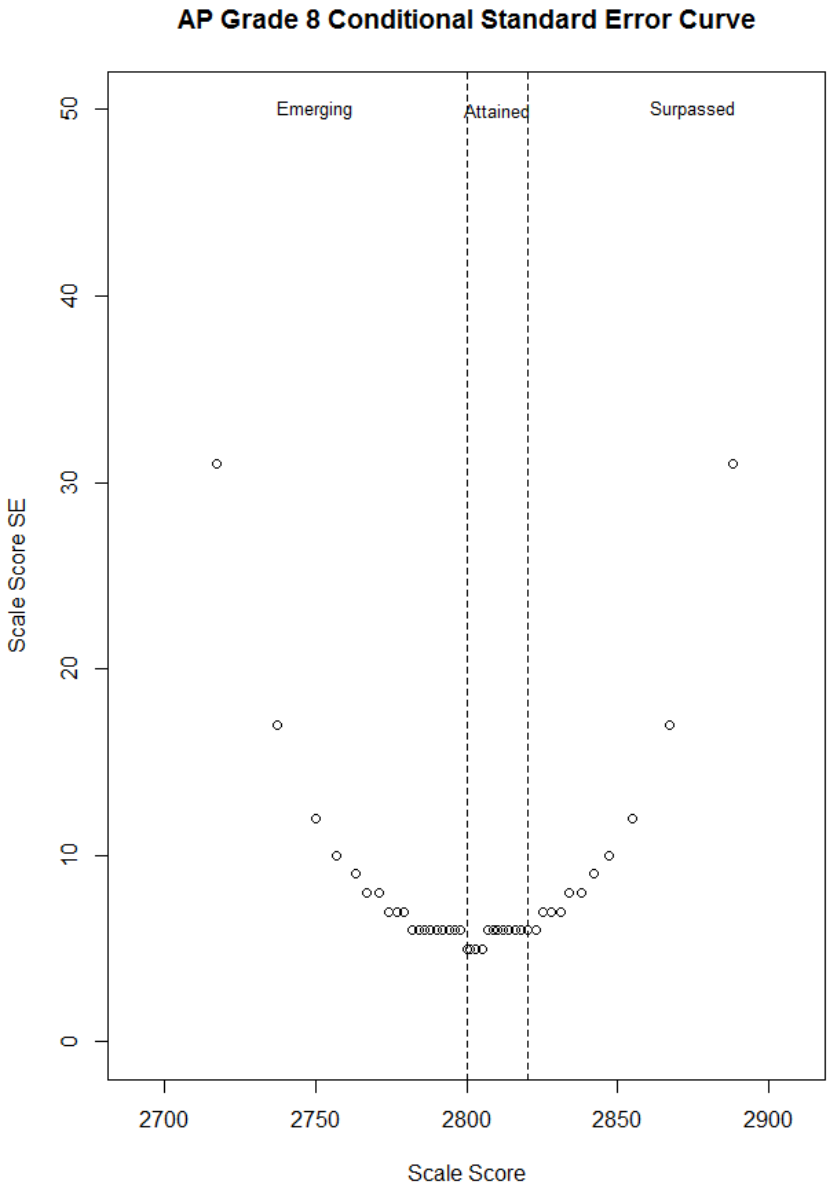

AP Grade 6 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
10
20
30
40
2500
2550
2600
2650
2700
Scale Score

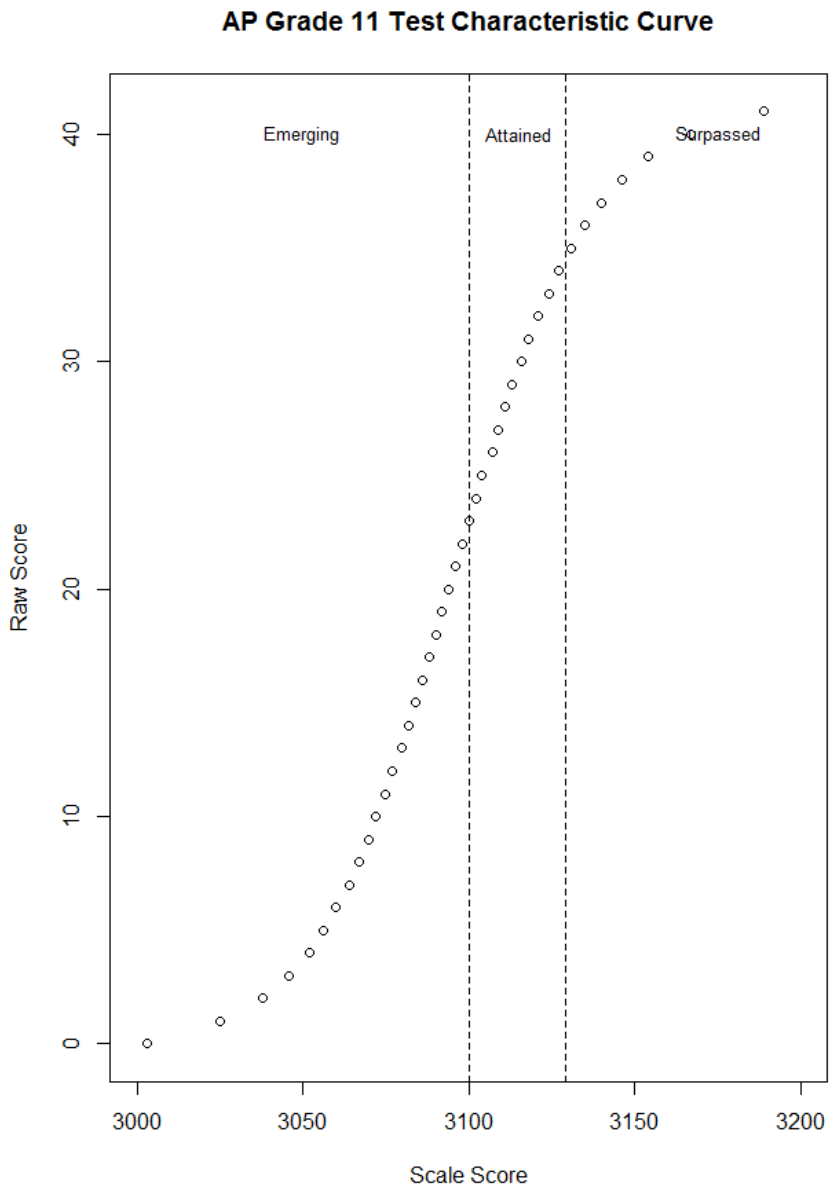

AP Grade 6 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
2500
2550
2600
2650
2700
Scale Score

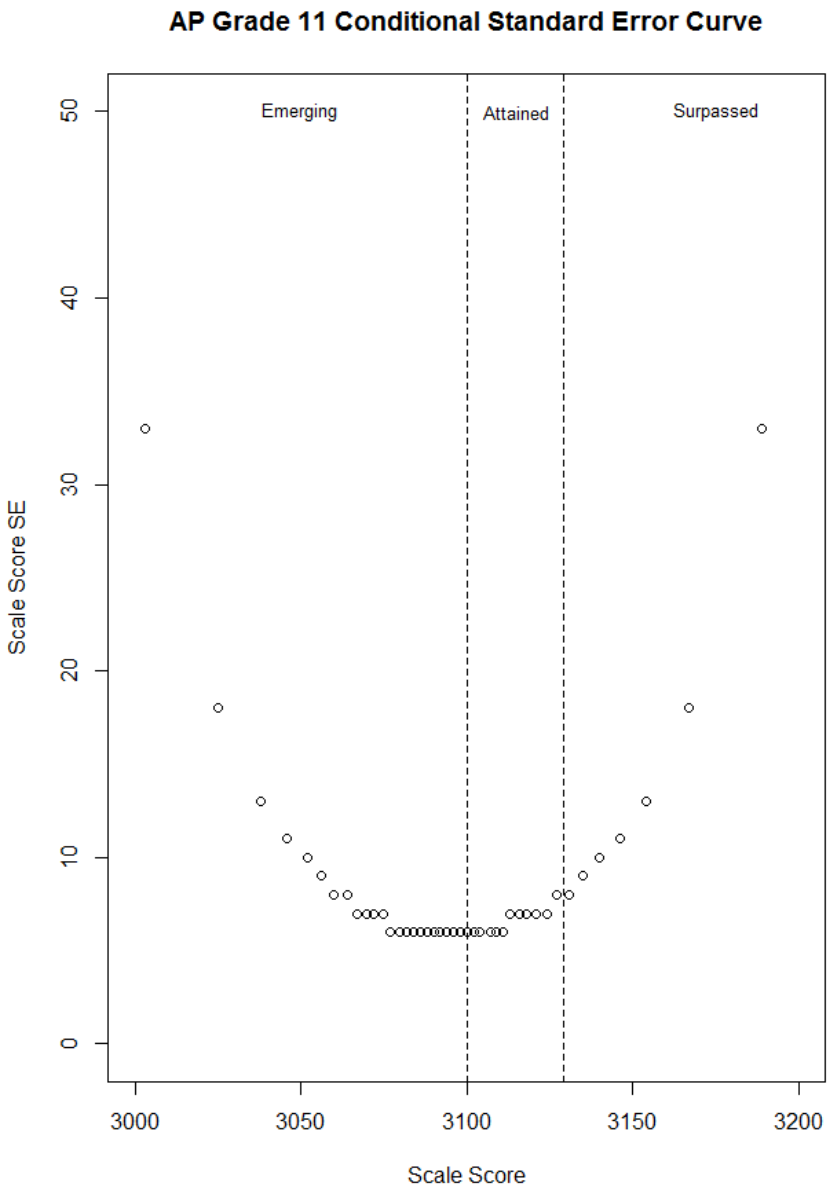

AP Grade 7 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
10
20
30
40
2600
2650
2700
2750
2800
Scale Score

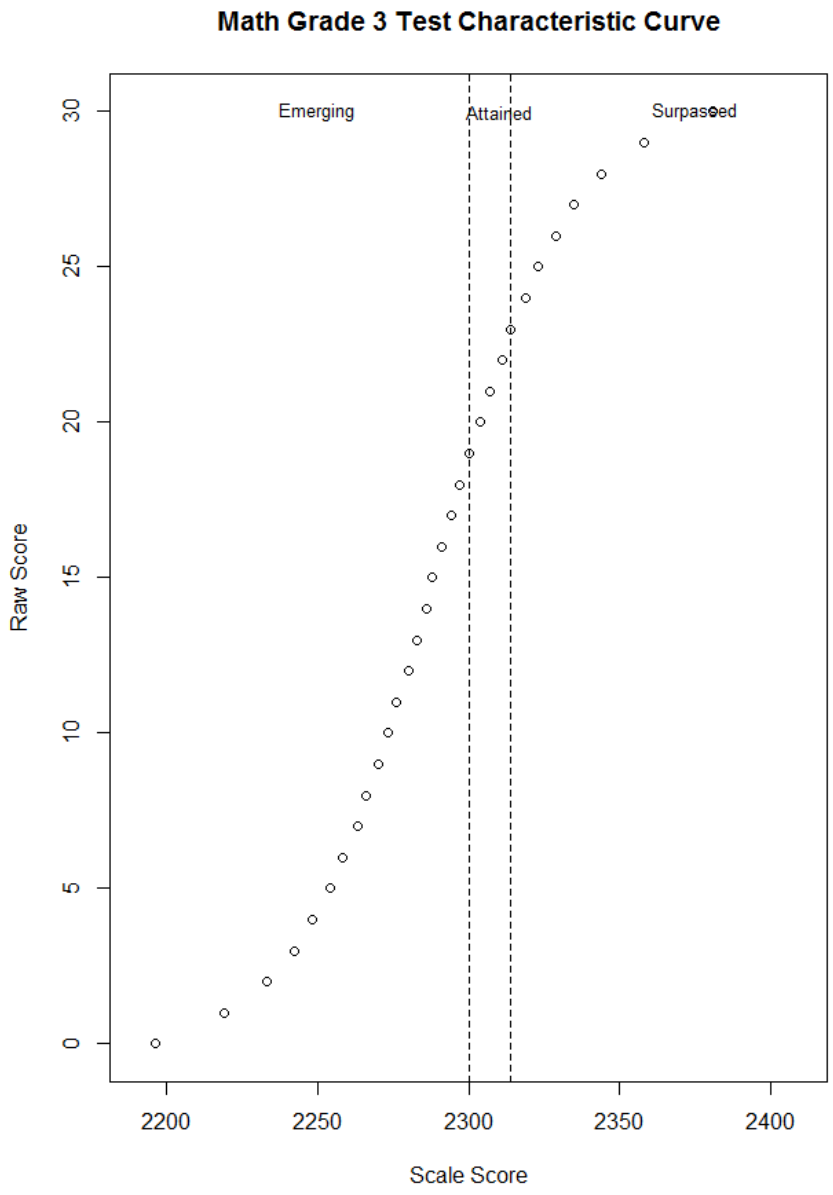

AP Grade 7 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
2600
2650
2700
2750
2800
Scale Score

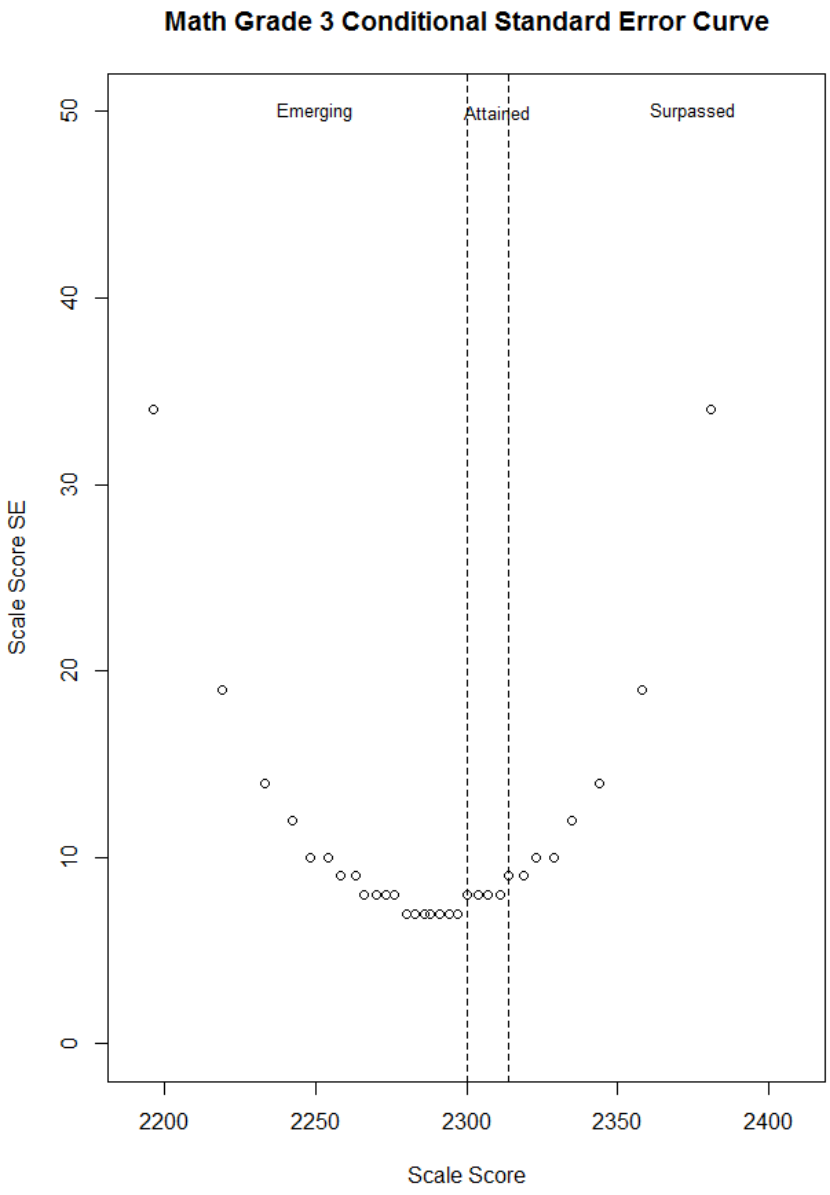

AP Grade 8 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
10
20
30
40
2700
2750
2800
2850
2900
Scale Score

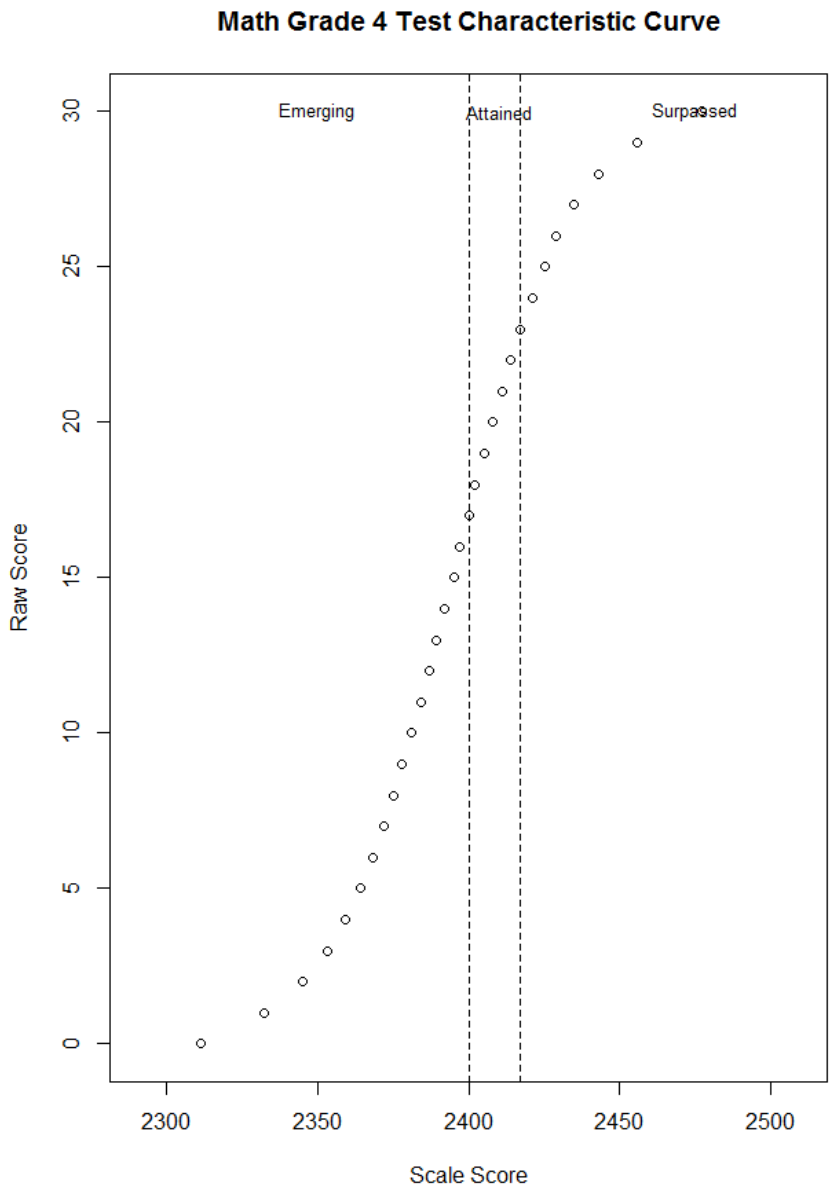

AP Grade 8 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
2700
2750
2800
2850
2900
Scale Score

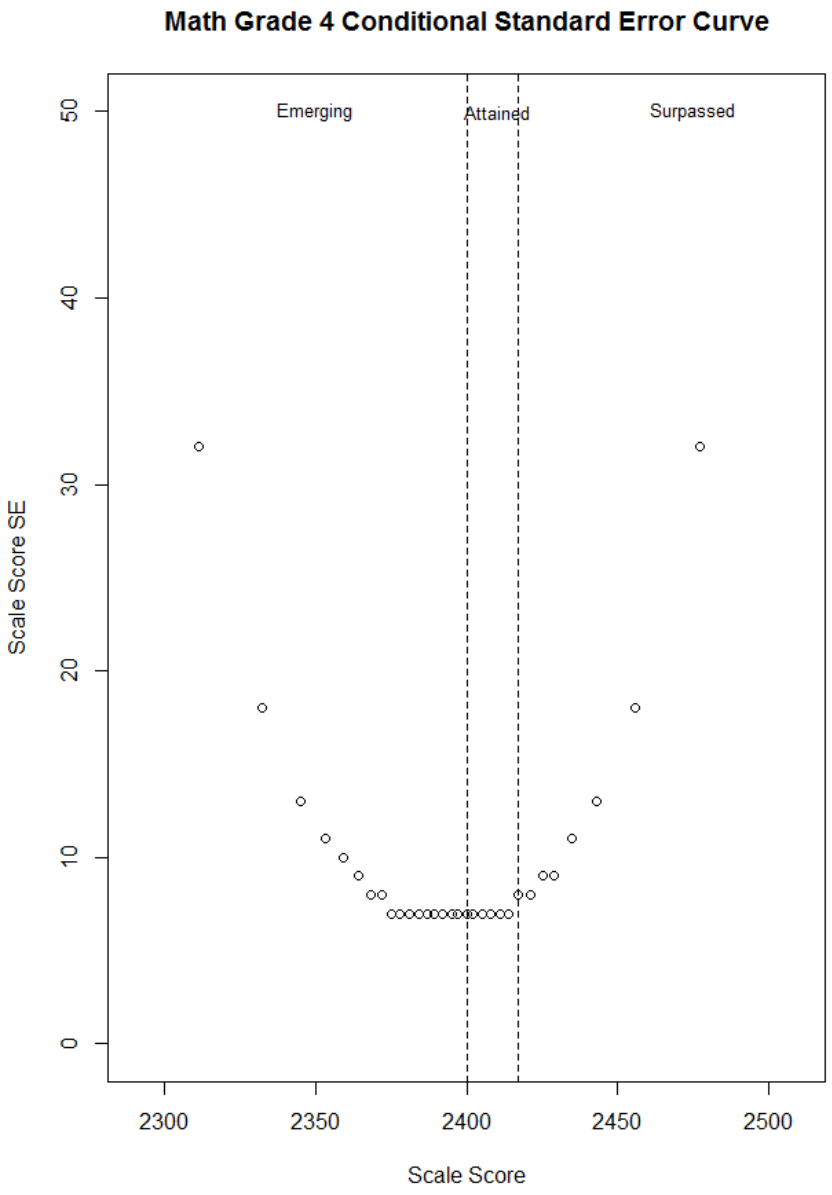

AP Grade 11 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
10
20
30
40
3000
3050
3100
3150
3200
Scale Score

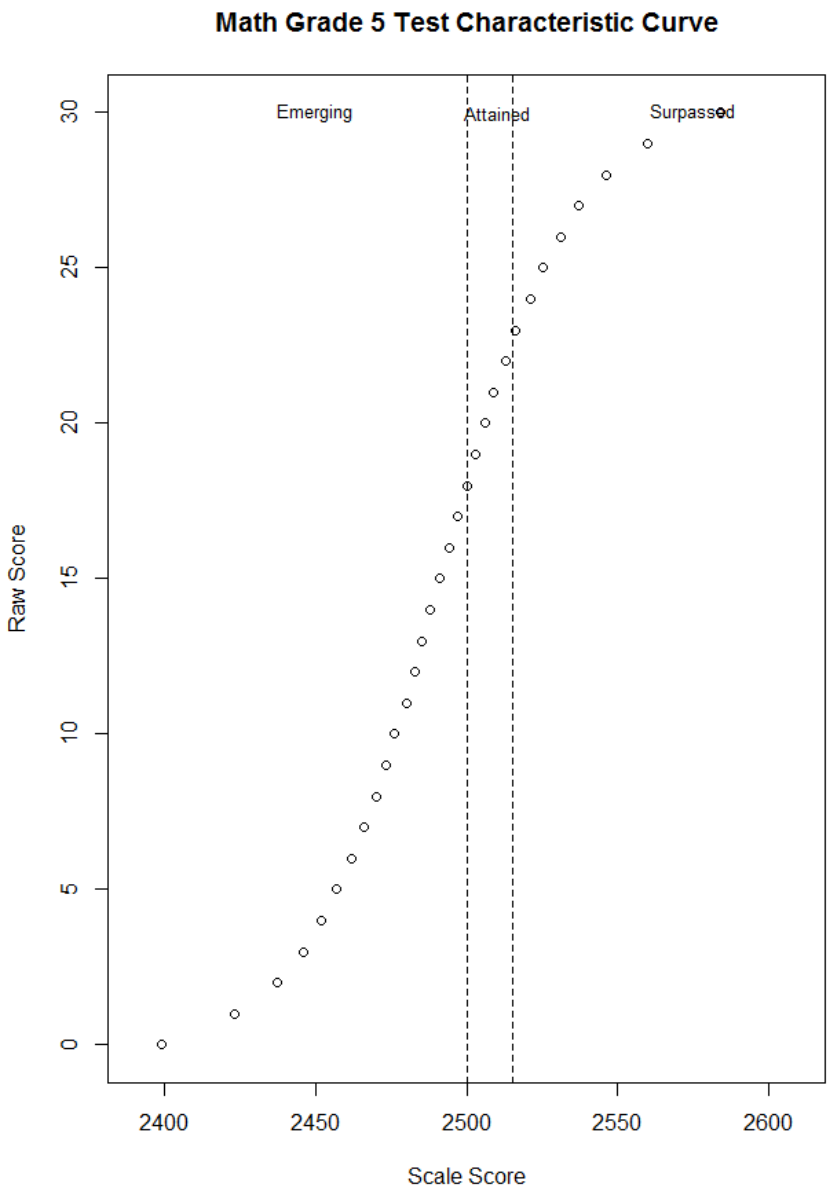

AP Grade 11 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
3000
3050
3100
3150
3200
Scale Score

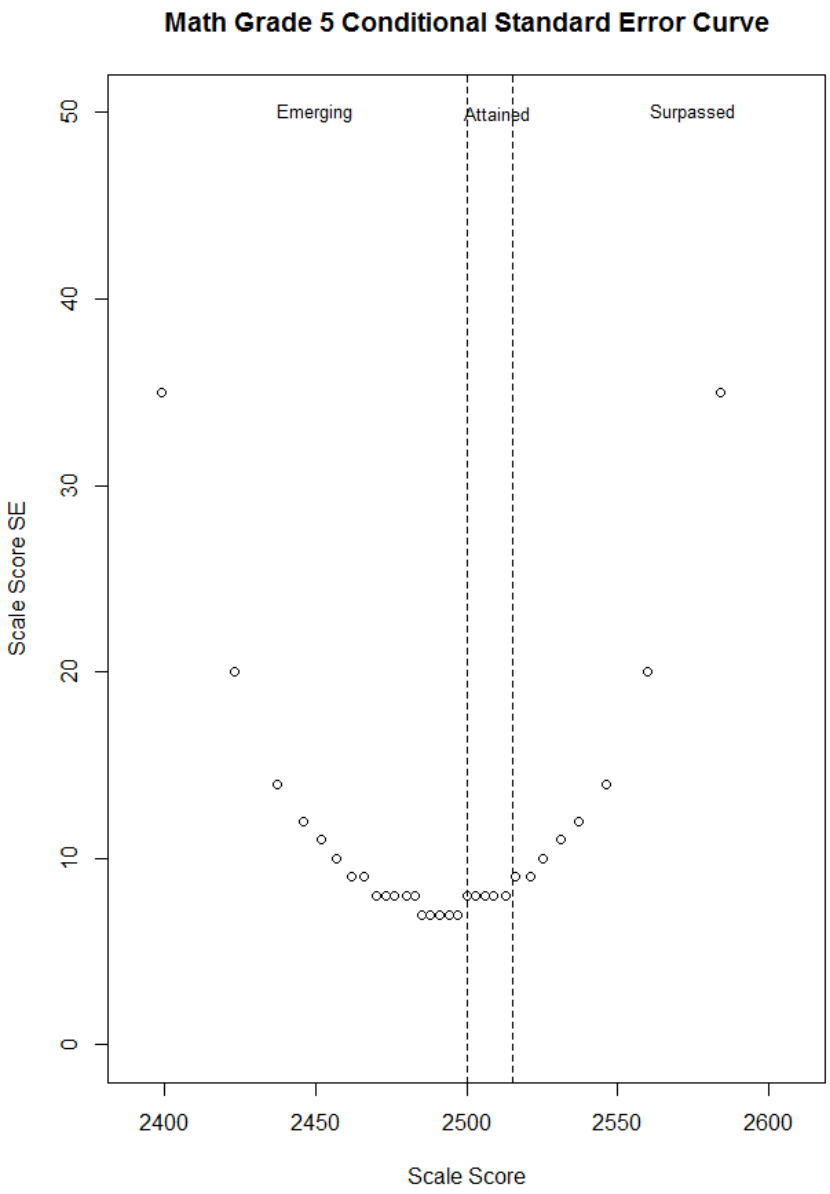

Math Grade 3 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
Scale Score
0
5
10
15
20
25
30
2200
2250
2300
2350
2400

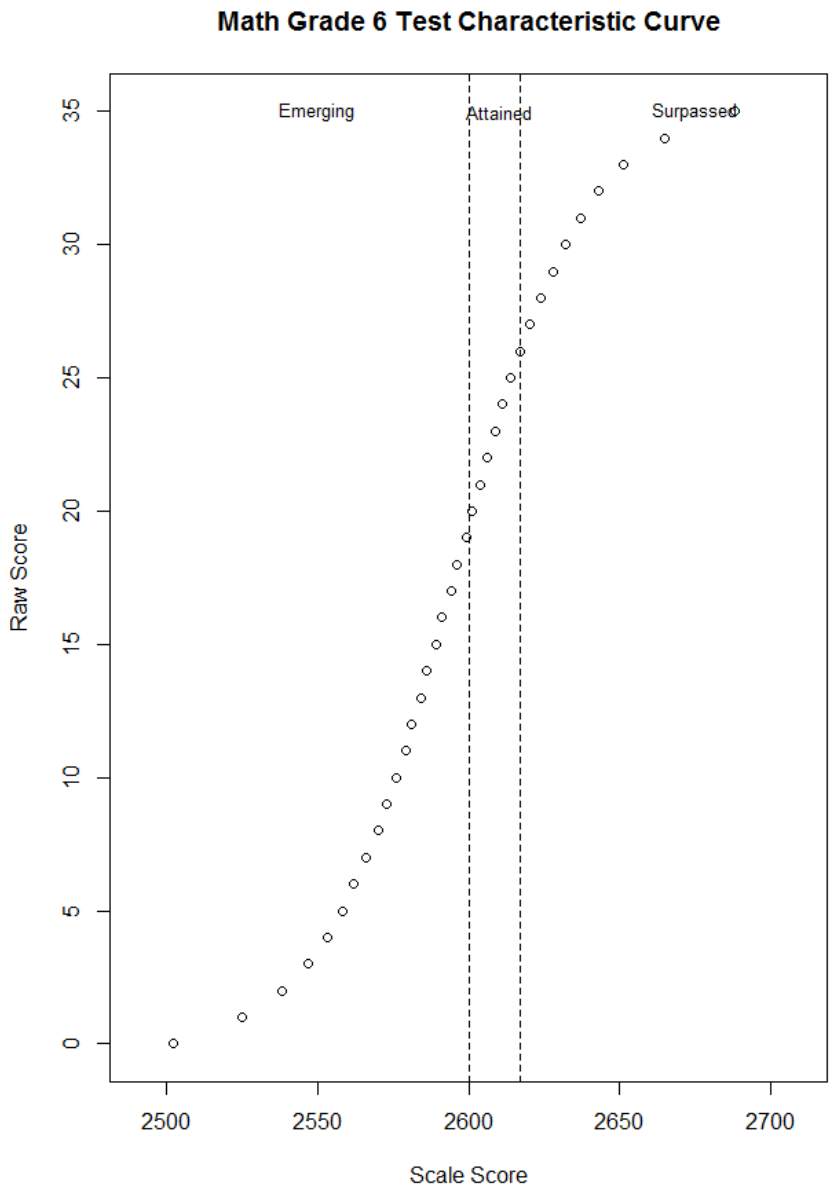

Math Grade 3 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
Scale Score
0
10
20
30
40
50
2200
2250
2300
2350
2400

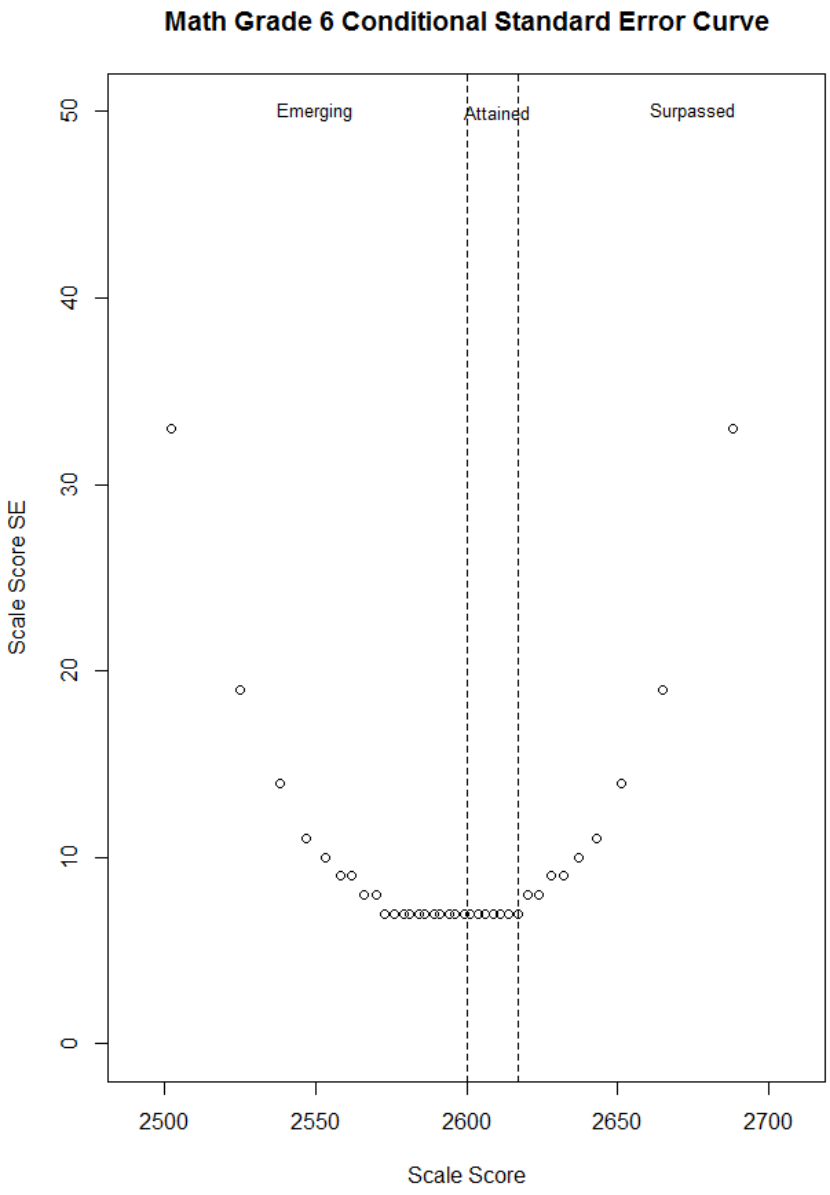

Math Grade 4 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
5
10
15
20
25
30
2300
2350
2400
2450
2500
Scale Score

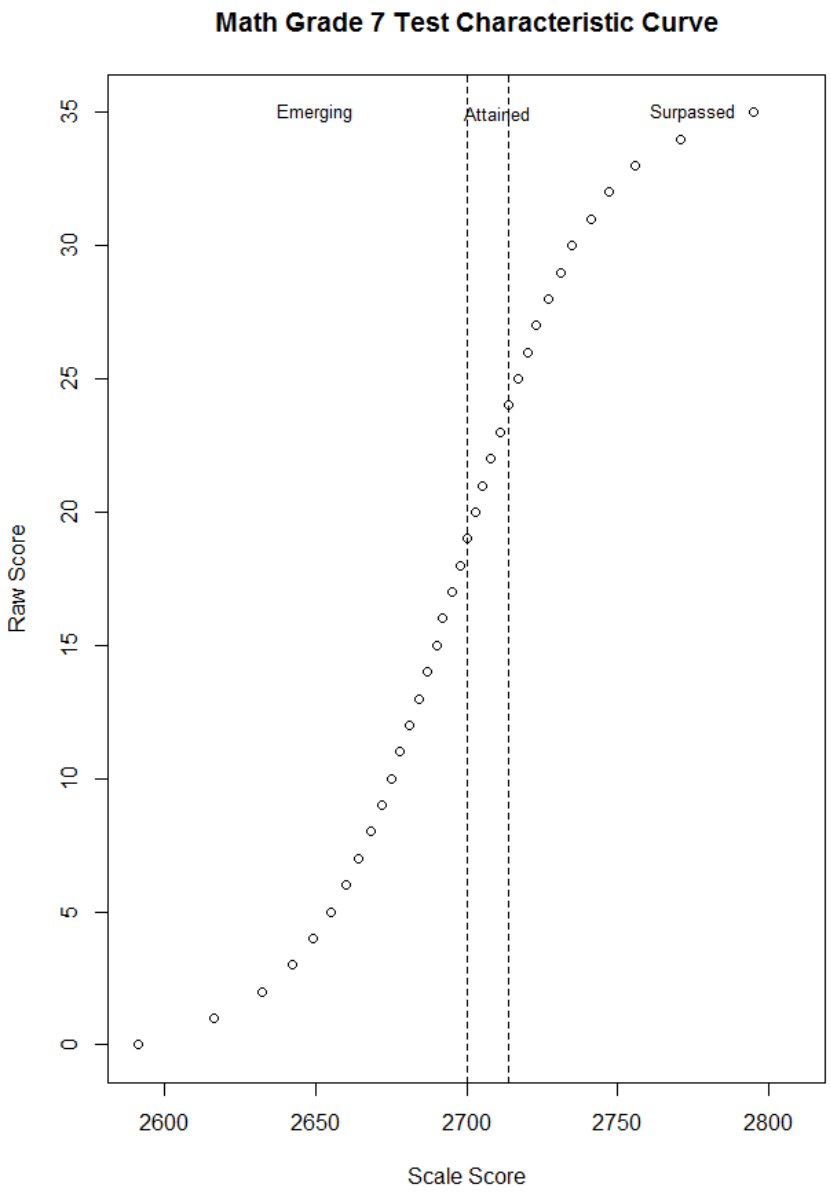

Math Grade 4 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
2300
2350
2400
2450
2500
Scale Score

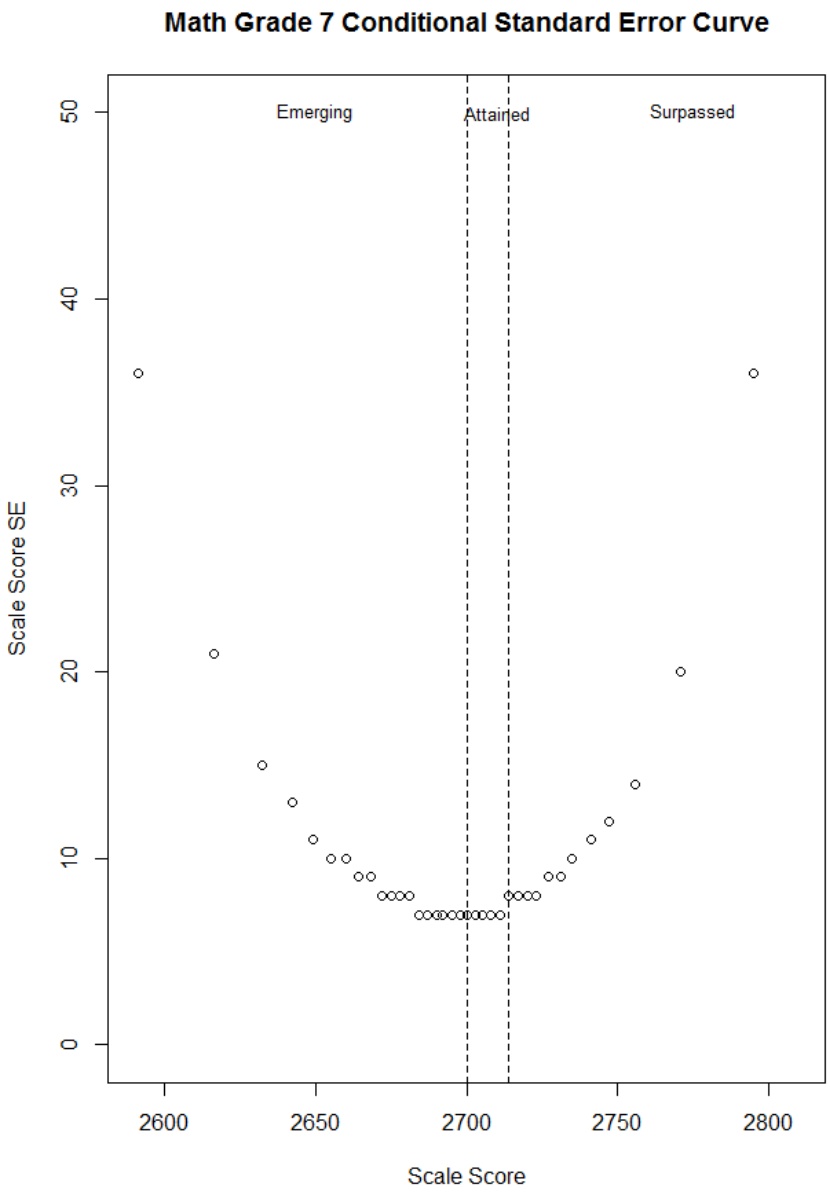

Math Grade 5 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
5
10
15
20
25
30
2400
2450
2500
2550
2600
Scale Score

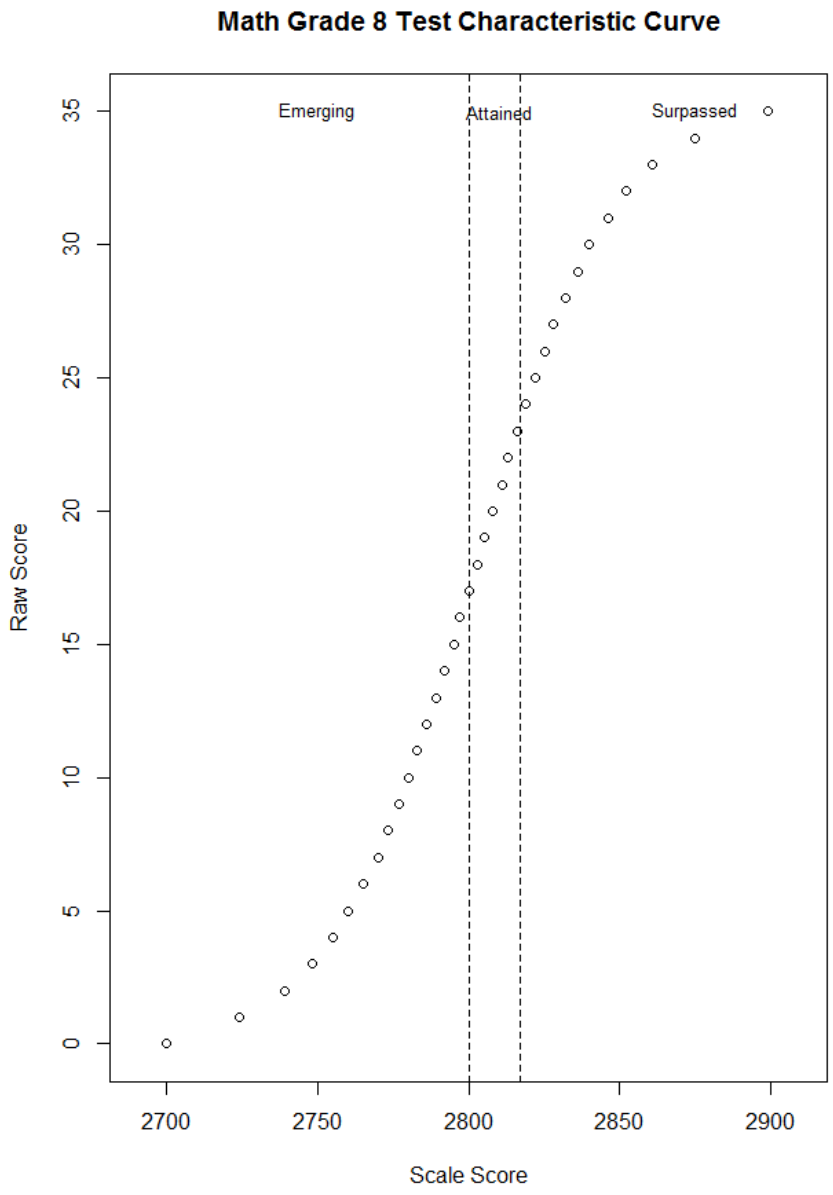

Math Grade 5 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
2400
2450
2500
2550
2600
Scale Score

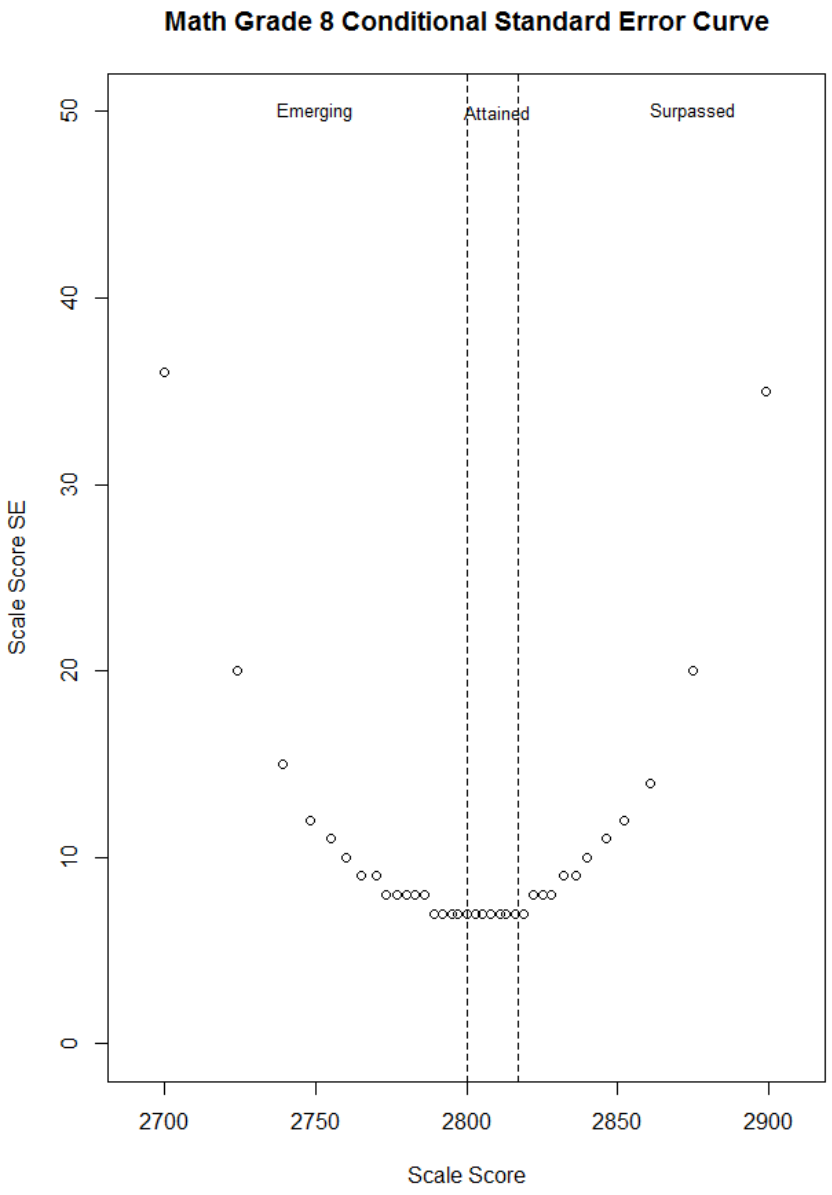

Math Grade 6 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
5
10
15
20
25
30
35
2500
2550
2600
2650
2700
Scale Score


Math Grade 6 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
2500
2550
2600
2650
2700
Scale Score

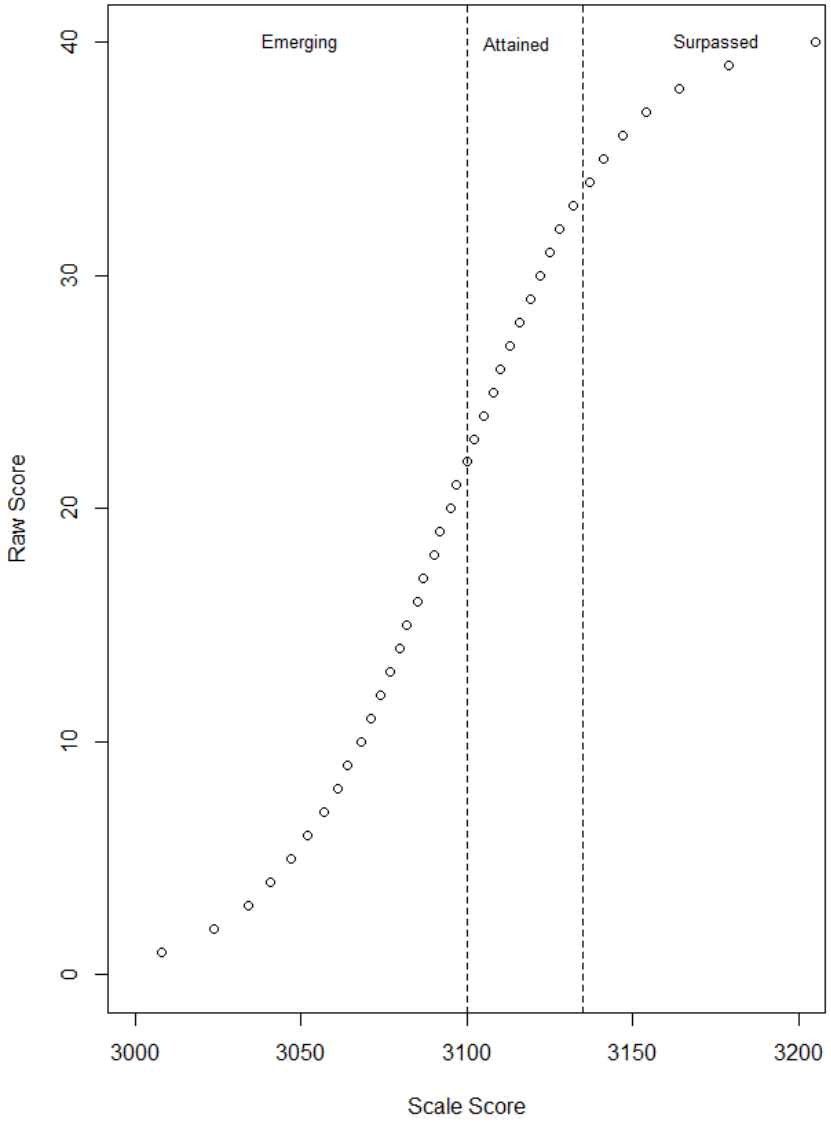

Math Grade 7 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
5
10
15
20
25
30
35
2600
2650
2700
2750
2800
Scale Score


Math Grade 7 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
2600
2650
2700
2750
2800
Scale Score

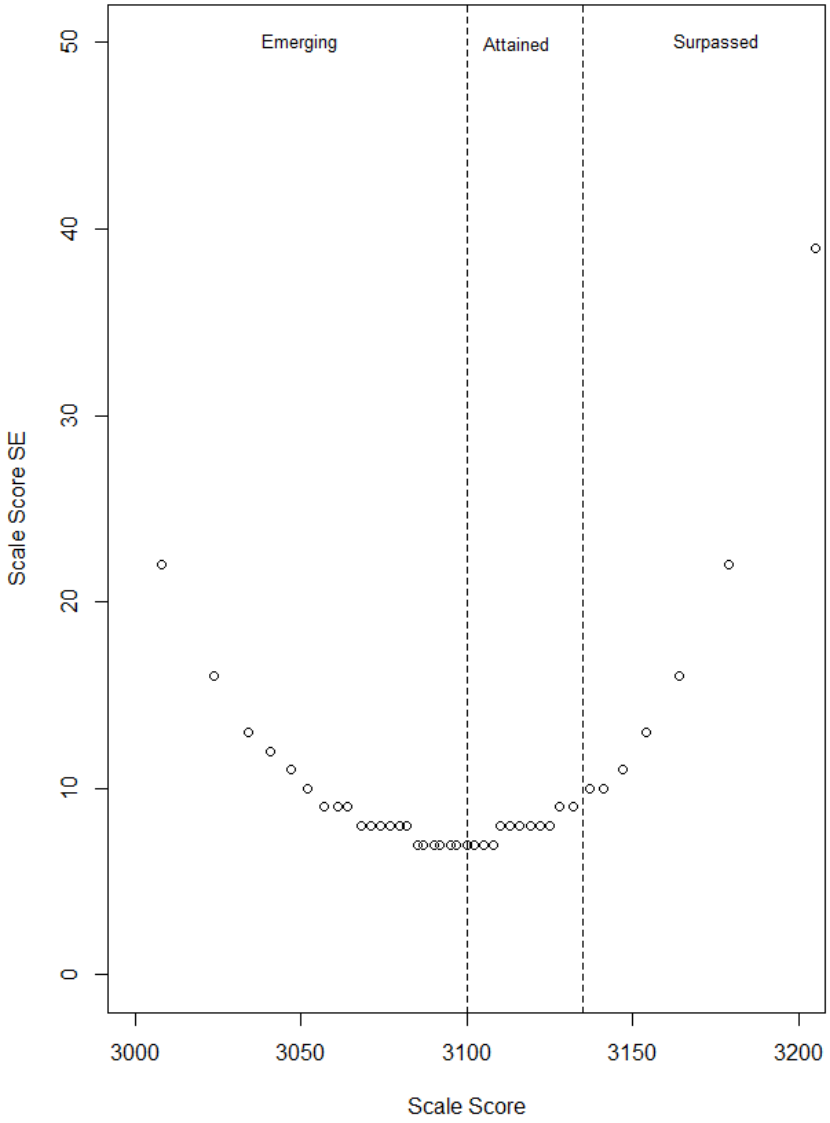

Math Grade 8 Test Characteristic Curve
Emerging
Attained
Surpassed
Raw Score
0
5
10
15
20
25
30
35
2700
2750
2800
2850
2900
Scale Score


Math Grade 8 Conditional Standard Error Curve
Emerging
Attained
Surpassed
Scale Score SE
0
10
20
30
40
50
2700
2750
2800
2850
2900
Scale Score

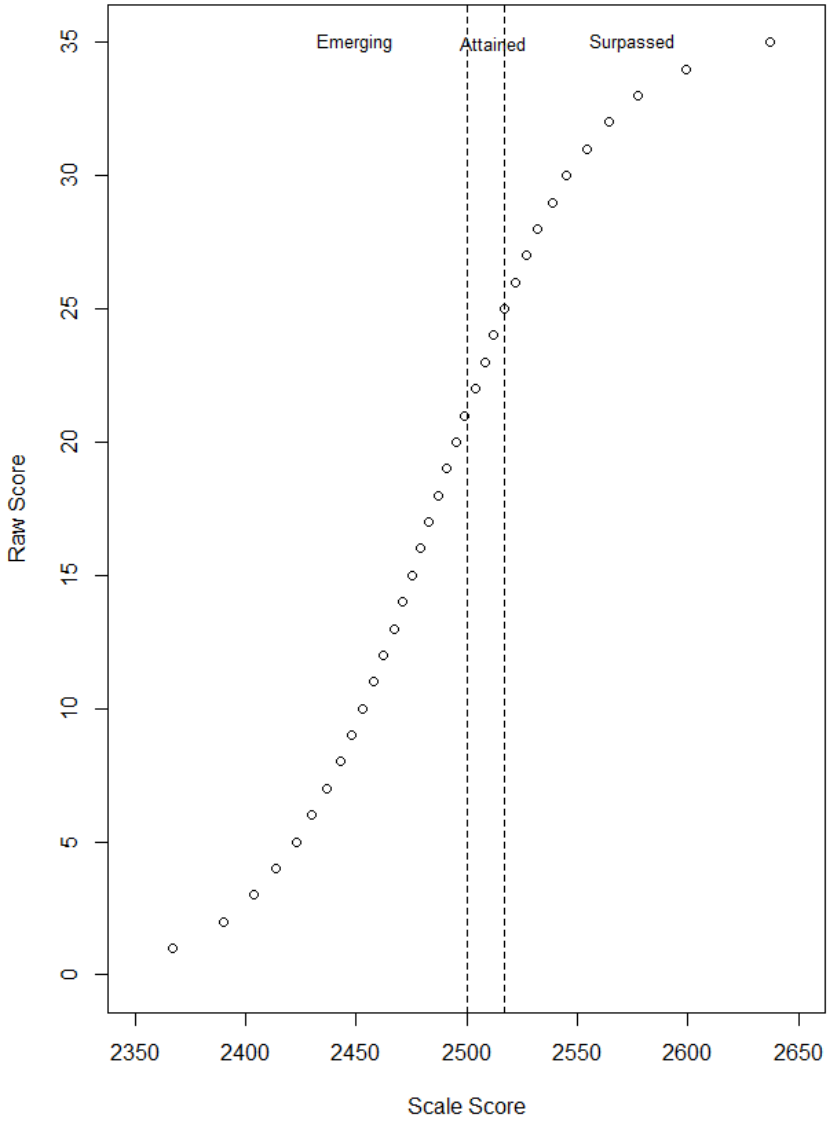
Math Grade 11 Test Characteristic Curve



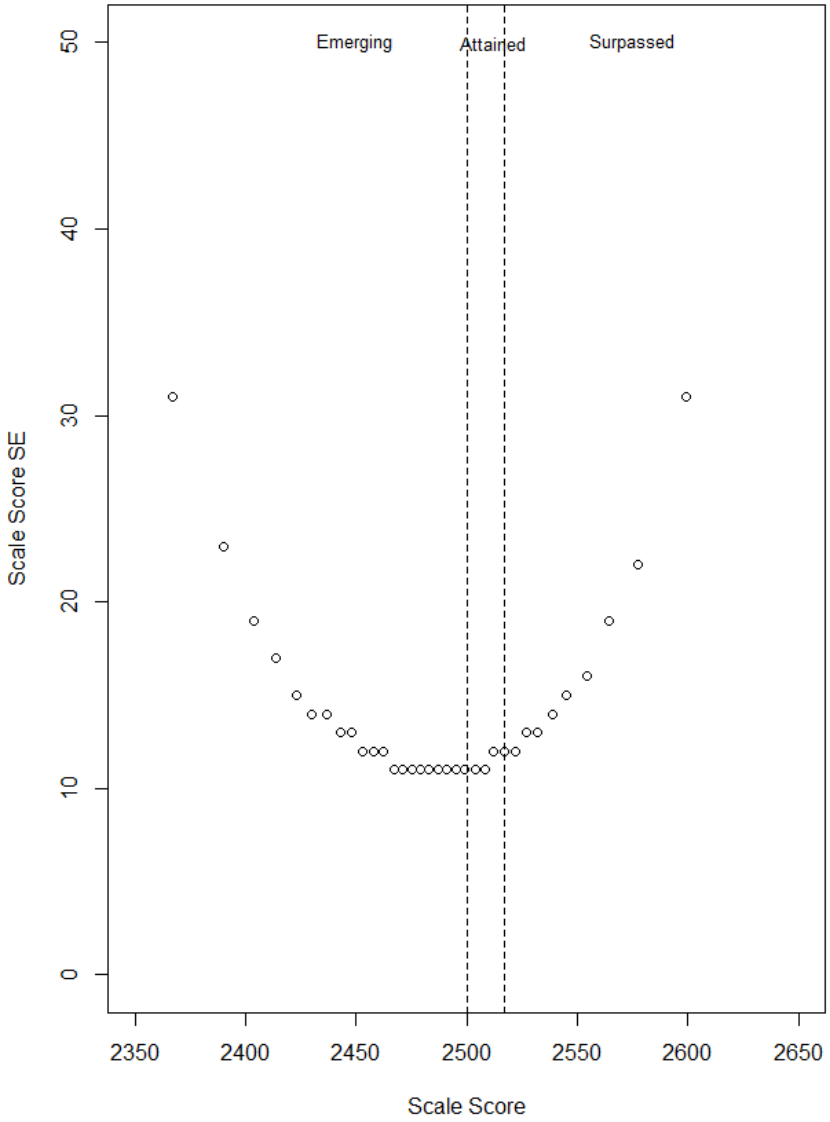
Math Grade 11 Conditional Standard Error Curve



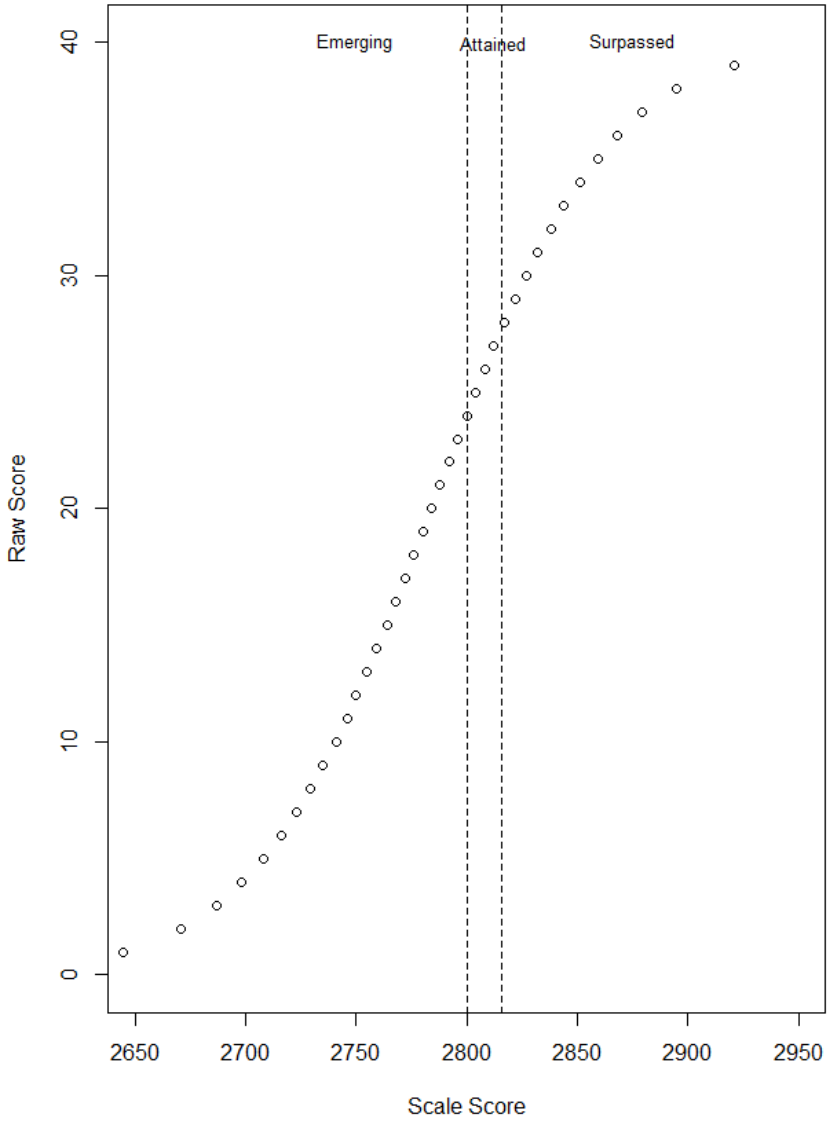
Science Grade 5 Test Characteristic Curve



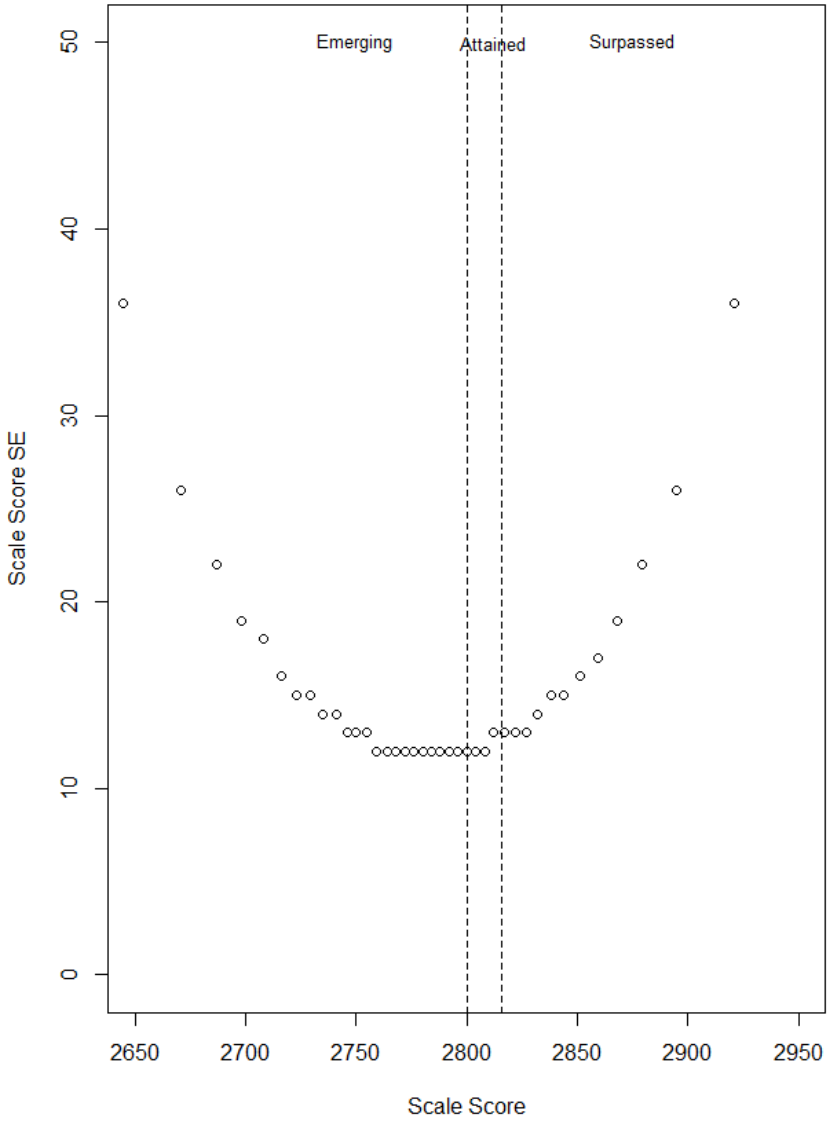
Science Grade 5 Conditional Standard Error Curve



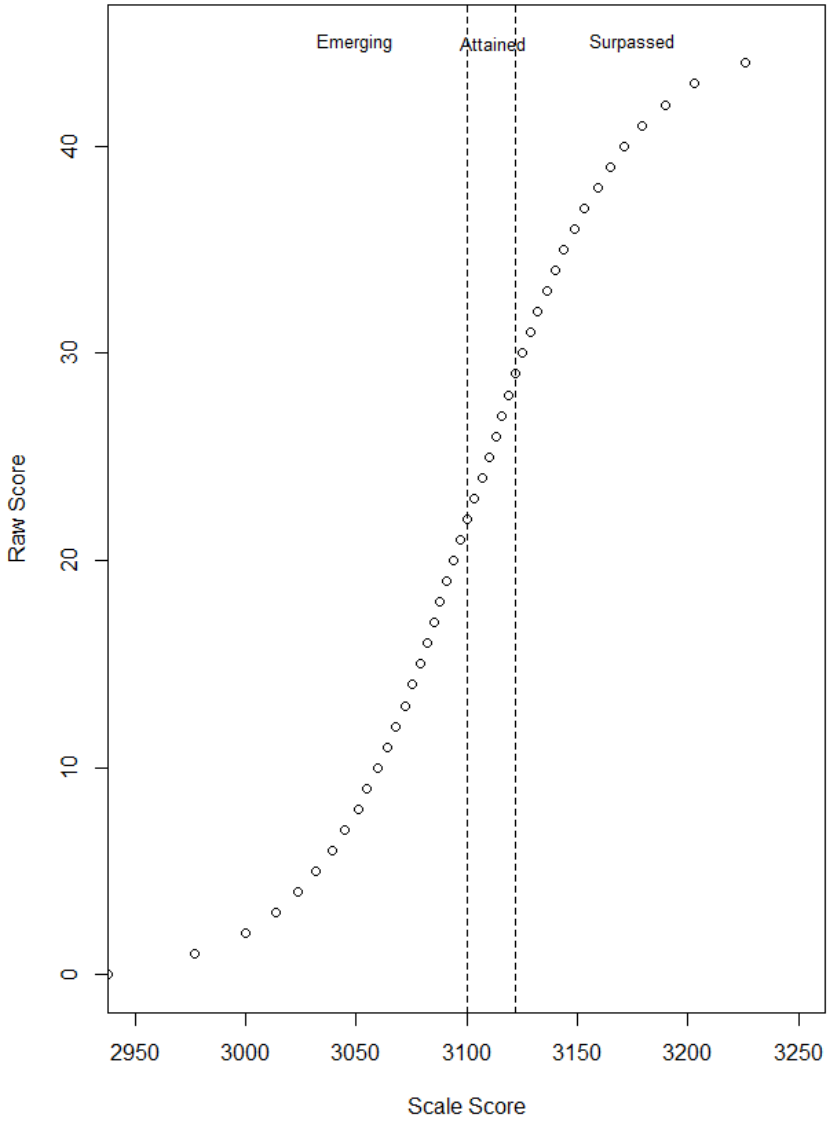
Science Grade 8 Test Characteristic Curve



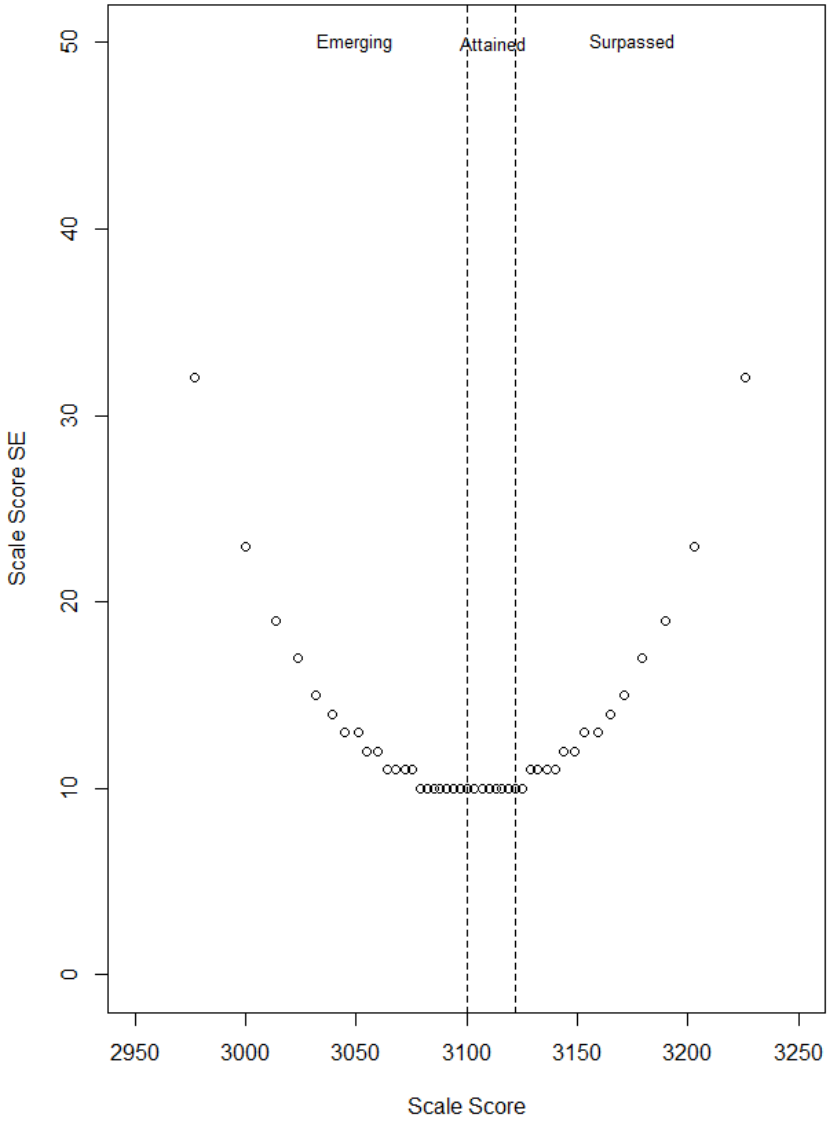
Science Grade 8 Conditional Standard Error Curve



Science Grade 11 Test Characteristic Curve



Science Grade 11 Conditional Standard Error Curve



Michigan Department of Education
Bureau of Assessment and Accountability
608 West Allegan Street
P.O. Box 30008
Lansing, MI 48909
(877) 560-8378
www.mi.gov/mi-access