Appendix C: Target Score Report

Psychometric Analysis Report for the
Michigan 3-7 English Language Arts (ELA) and Mathematics Assessment Target Reporting
2018-2019

November 2019

DRC Psychometric Services Michigan Project Team
Data Recognition Corporation

Table of Contents

List of Tables

Introduction

The assessment target score report is designed to report a group of students' (e.g., at the grade, school, teacher, and/or district levels) relative strength and weakness at the assessment target level. It is for aggregate level reports only.

Unlike the performance categories provided at the total test and claim levels, these strengths and weakness do not imply proficiency. Instead, they show how a group of students' performance is distributed across the content target relatively to their overall performance. For example, a group of students may have performed very well on a subject, but performed lower on a target. Thus, performance level code of C not necessarily imply a lack of proficiency, but that these students' performance on that target was lower than their performance across other targets put together. It can be concluded that the students performed lower than expected on that target.

Assessment target score report should serve as a starting point in an overall investigation of students' strengths and weaknesses and constitutes only one of many sources of evidence that should be used in evaluating student performance.

This was conducted for the English Language Arts (ELA) and Mathematics M-STEP assessments.

Methodology

Item response theory (IRT) based residual analysis can be used to conduct analyses for the assessment target score report. The residual is the difference between the observed score and expected score at the item level. The observed score is the score (e.g., 0 to 3) a student submitted for each item. The expected score is derived using the 2 parameter logistic (2PL) model for dichotomously scored items and generalized partial credit model (GPCM) for polytomously scored items.

The expected score for a multiple-choice item (MC, one point item) was computed using the two-parameter logistic (2PL) model as shown below in equation 1.

$$P_i\left(x_i = 1|\theta, a_i, b_i\right) = \frac{\exp Da_i(\theta - b_i)}{1+\exp Da_i(\theta - b_i)}, \quad (1)$$

where a_i is item discrimination parameter and b_i is item difficulty for item *i*, and p_i is the probability of the item getting correct given the observed overall ability estimate, θ, and *D* is 1.7. The expected score for a constructed response (CR) item, the observed overall ability estimate, θ, was computed with generalized partial credit mode (equation 2).

$$P_{ik}\left(x_i|\theta, a_i, b_{i0}, b_{i1}, \ldots, b_{iK_i}\right) = P_{ik}\left(\theta\right) = \frac{\exp Da_i \sum_{k=0}^{x_i}(\theta - b_{ik})}{\sum_{r=0}^{K_i}\left[\exp Da_i \sum_{k=0}^{r}(\theta - b_{ik})\right]},$$

where $\sum_{k=0}^{0}(\theta - b_{ik}) \equiv 0.$ (2)

Equation (2) computes the probability of obtaining the score of $0 \leq x_i \leq$ $_i$ on CR item . The item discrimination parameter is a_i, and b_i is the category intersection parameter (in SBAC scoring specification, it is referred to step parameters). Equation (1) is a special case of equation (2) with $_i =$ 1. This means that the computation of probability can be completed for both 2PL MC and CR items using equation (2).

For all items, the residual, R_i, is found by using equation (3),

$$R_i = O_i - P_i\left(\theta\right) \tag{3}$$

where O_i is the observed score for item *i* and $P_i\left(\theta\right)$ is the expected score for item *i.*

Once the individual residuals were calculated, the weighted average of the residuals were calculated for each assessment target meeting the reporting criteria (see the Reporting Criteria Section below for more details) criteria using equation (4).

$$\bar{R}_{target} = \frac{1}{\sum_{i=1} w_i} \sum_{i=1} R_i \tag{4}$$

where R_i is the residual for item *i* and w_i is the weight associated with item *i* that accounts for the number of score points for that item.

Reporting Criteria

Target assessment results were reported for both ELA and Mathematics. Table 1 provides the claim and target level for which target assessment results were reported. Once the average residual for each assessment target was computed, a flagging criterion of +/- 0.05 was used to indicate the assessment target level performance. Table 2 provides a description of the performance levels.

Table 1: Assessment targets are listed below by content area.

ELA	
Claim	**Assessment Target**
Claim 1	1
	2
	3
	4
	5
	6
	7
	8
	9
	10
	11
	12
	13
	14
Claim 2	1
	2
	3
	4
	6
	7
	8
	9
Claim 3	4
Claim 4	2
	3
	4

Math	
Claim	**Assessment Target**
Claim 1	A
	B
	C
	D
	E
	F
	G
	H
	I
	J
	K
	L
Claim 2	A
	B
	C
	D
Claim 3	A
	B
	C
	D
	E
	F
Claim 4	A
	B
	C
	D
	E
	F

Table 2: Performance level descriptions.

PL Code	Target Level	Description
A	Better than performance on the test as a whole	This target is a relative strength. The group of students performed better on items from this target than they did on the rest of the test, as a whole.
B	Similar to performance on the test as a whole	This target is neither a relative strength nor a relative weakness. The group of students performed about as well on items from this target as they did on the rest of the test, as a whole.
C	Worse than performance on the test as a whole	This target is a relative weakness. The group of students did not perform as well on items from this target as they did on the rest of the test, as a whole.
--	Insufficient Information	Not enough information is available to determine whether this target is a relative strength or weakness.

Additionally, since the M-Step administration was a CAT, the number of items presented in each assessment target varied for each administration. Thus, reporting criteria was used to ensure that a specified number of unique items were presented in order the assessment target results to be provided. The criteria used is listed below:

- Number of unique students per target: n=15
- Number of unique items per target: n=3
- Number of responses per target: n=25
- Use 0.05 criterion on the rescaled residual scale

Exclusions

It should be noted that some students were excluded from the target reporting analysis. Students who were Force Submit or scored at the lowest and highest obtainable scale score (LOSS and HOSS) were excluded from the analysis. Additionally, students with invalid tests and home schooled students were excluded.

Results

Aggregate results were provided to MDE by State, ISD, District, and Building. Private school students were only included in the building level aggregate results. Table 3 shows N count used at the state level and aggregates for all students and without and with disabilities (AED and SWD) and Tables 4 and 5 provide the state level results for ELA and Mathematics. Note that the PL codes in Tables 3 and 4 correspond to those found in Table 2.

Table 3: Valid student counts at the state level by content area and grade.

Content Area	Grade	N	N AED	N SWD
ELA	**3**	100,038	88,154	11,884
ELA	**4**	101,433	89,324	12,109
ELA	**5**	104,462	91,929	12,533
ELA	**6**	108,163	95,855	12,308
ELA	**7**	107,607	95,763	11,844
Mathematics	**3**	99,844	88,033	11,811
Mathematics	**4**	101,675	89,610	12,065
Mathematics	**5**	104,492	92,026	12,466
Mathematics	**6**	107,855	95,873	11,982
Mathematics	**7**	107,719	96,209	11,510

Table 4a: State level aggregate results for ELA.

Claim	Target	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7	
		Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code
1	**1**	71288	B	98166	B	76495	B	93157	B	106992	B
	2	97773	B	88687	B	83653	B	84217	C	52699	B
	3	89635	B	85448	B	89712	B	101487	B	89716	B
	4	79978	B	71164	B	93291	B	80593	C	82631	D
	5	49096	B	23499	B	18134	B	103729	C	67504	D
	6	23941	B	67737	B	86464	B	84311	B	71379	B
	7	88817	B	85613	B	71262	B	78010	B	76969	B
	8	84953	B	93611	B	93468	B	95794	B	95759	B
	9	83270	B	94046	B	63617	B	82246	B	105815	B
	10	86764	B	95730	B	93437	B	82851	B	84669	A
	11	72135	B	88569	B	86965	B	107489	B	99184	B
	12	70032	B	48548	B	66863	B	13510	B	64504	A
	13	81589	A	54698	B	81077	B	61770	B	66777	B
	14	71269	B	67305	C	91176	B	50096	C	63389	A
2	**1**	96932	B	88494	B	100784	B	105540	B	73809	B
	2										
	3	91198	B	95216	B	96013	C	93919	C	103916	C
	4	100038	A	101433	A	103829	B	108163	B	107607	B
	6	65382	C	74385	B	78736	C	82900	B	102654	C
	8	100038	B	101433	B	104462	B	108163	B	107607	B
	9	100038	B	101433	B	104462	B	108163	B	107607	B
3	**4**	100038	B	101433	B	104462	B	108163	B	107607	B
4	**2**	100036	B	101430	B	103580	B	103514	B	97795	A
	3	75531	B	76532	B	101954	B	107976	B	106511	B
	4	84917	B	97243	B	99898	B	99057	B	106168	B

Table 4b: State level aggregate results for ELA for Students without Disabilities

Claim	Target	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7	
		Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code
1	**1**	63974	B	86577	B	65960	B	82663	B	95341	B
	2	86380	B	77266	B	74773	B	75765	C	48840	B
	3	78372	B	74731	B	78059	B	89464	B	78419	B
	4	71294	B	64399	B	83523	B	73157	C	74227	D
	5	44833	B	21553	B	16737	B	91637	C	62880	D
	6	19936	B	58827	B	75650	B	73505	B	63406	B
	7	77719	B	75020	B	62475	B	68776	B	67286	B
	8	75798	B	82862	B	82475	B	84716	B	84347	B
	9	73006	B	82900	B	54138	B	74264	B	94027	B
	10	75553	B	84026	A	81580	B	72818	B	74730	A
	11	65145	B	79037	B	77708	B	95374	B	89292	B
	12	61987	B	41687	B	61613	B	12604	B	57812	A
	13	71353	B	47469	B	69768	B	53354	B	60133	B
	14	62050	B	57234	C	80407	B	43393	B	54529	A
2	**1**	85614	B	77037	C	88419	B	93424	B	65061	B
	2										
	3	80199	B	83998	B	84033	C	82873	C	92282	C
	4	88154	A	89324	A	91468	B	95855	B	95763	B
	6	57201	C	66286	C	70662	C	74268	B	91896	C
	8	88154	B	89324	B	91929	B	95855	B	95763	B
	9	88154	B	89324	B	91929	B	95855	B	95763	B
3	**4**	88154	B	89324	B	91929	B	95855	B	95763	B
4	**2**	88152	B	89321	B	91115	B	92009	B	86723	A
	3	65246	B	65748	B	89562	B	95670	B	94692	B
	4	75419	B	86128	B	88692	B	87652	B	94722	B

Table 4c: State level aggregate results for ELA for Students with Disabilities

Claim	Target	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7	
		Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code
1	**1**	7314	B	11589	B	10535	B	10494	A	11651	B
	2	11393	C	11421	B	8880	B	8452	B	3859	B
	3	11263	B	10717	B	11653	B	12023	B	11297	A
	4	8684	B	6765	B	9768	B	7436	B	8404	D
	5	4263	B	1946	B	1397	B	12092	B	4624	D
	6	4005	B	8910	B	10814	B	10806	B	7973	B
	7	11098	B	10593	B	8787	B	9234	B	9683	B
	8	9155	B	10749	B	10993	A	11078	A	11412	B
	9	10264	B	11146	B	9479	A	7982	B	11788	B
	10	11211	B	11704	B	11857	A	10033	B	9939	A
	11	6990	B	9532	B	9257	B	12115	B	9892	B
	12	8045	B	6861	B	5250	B	906	B	6692	A
	13	10236	A	7229	B	11309	B	8416	C	6644	B
	14	9219	B	10071	B	10769	B	6703	C	8860	A
2	**1**	11318	B	11457	B	12365	B	12116	B	8748	B
	2										
	3	10999	B	11218	B	11980	C	11046	C	11634	C
	4	11884	B	12109	A	12361	B	12308	B	11844	C
	6	8181	B	8099	B	8074	C	8632	B	10758	B
	8	11884	B	12109	B	12533	B	12308	B	11844	B
	9	11884	C	12109	B	12533	C	12308	B	11844	C
3	**4**	11884	B	12109	B	12533	B	12308	B	11844	B
4	**2**	11884	B	12109	B	12465	B	11505	B	11072	A
	3	10285	B	10784	B	12392	B	12306	B	11819	B
	4	9498	B	11115	B	11206	B	11405	B	11446	B

Table 5a: State level aggregate results for Mathematics.

Claim	Target	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7	
		Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code
1	A	99844	B	78065	B	23194	B	107855	B	107320	B
	B	34951	B	57596	C	2053	C	81614	B	105379	B
	C	85082	B	22047	B	101021	B	96072	B	106488	B
	D	90266	B	101675	B	77383	B	107855	A	107263	B
	E	99007	B	55439	B	104065	B	73827	B	98876	B
	F	84070	B	96357	B	104492	B	107685	A	81507	B
	G	84948	B	101675	A	71947	B	101662	B	68126	B
	H	99844	B	101675	B	67493	B	84746	B	39045	C
	I	40485	A	99791	B	83354	B	24542	B	71004	B
	J	31536	B	22032	B	103119	B	52826	B		
	K	26105	A	49958	A						
	L			101664	B						
2	A	99844	B	101675	B	104492	B	107855	B	107719	B
	B	76079	B	16874	B	37639	B	47189	B	51367	B
	C	59492	B	95583	B	49671	B	57478	B	62831	B
	D	18396	B	23179	B	81737	B	71888	B	68920	B
3	A	88501	B	96750	B	94311	B	106658	B	101490	B
	B	92861	B	86877	B	62839	B	42672	B	23632	D
	C	69775	B	73649	B	76039	B	75254	B	75146	B
	D	90492	B	82748	B	91591	B	68046	B	91063	B
	E	87055	B	95911	B	100749	B	107768	B	107719	B
	F	74627	B	82147	B	81563	B	78172	B	62402	B
	G							22372	B	45406	B
4	A	61592	B	92218	B	84359	B	103732	B	90747	B
	B	2144	A	21706	B	45911	C	16676	B	12853	B
	C	56244	B	60958	B	51546	B	30035	B	53678	B
	D	87478	B	56348	B	77243	B	44308	B	71697	B
	E	97700	B	79969	B	58581	B	91183	B	94872	B
	F	43600	B	40728	B	52946	B	77820	B	54041	B

Table 5b: State level aggregate results for Mathematics for Students without Disabilities.

Claim	Target	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7	
		Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code
1	A	88033	B	67106	B	17407	B	95873	B	95833	B
	B	29934	B	48376	C	1549	C	74493	B	94167	B
	C	74008	B	20221	B	89416	B	84431	B	95103	B
	D	79037	B	89610	B	66355	B	95873	A	95895	B
	E	87489	B	49906	B	91705	B	64903	B	87767	B
	F	75077	B	86182	B	92026	B	95707	A	73455	B
	G	76259	B	89610	A	63339	B	90142	B	59238	B
	H	88033	B	89610	B	61989	B	78001	B	35880	C
	I	36478	A	88165	B	72655	B	20832	B	64358	B
	J	26744	B	21013	B	90692	B	44556	B		
	K	20312	A	42683	A	31267	B				
	L			89610	B						
2	A	88033	B	89610	B	92026	B	95873	B	96209	B
	B	66370	B	16037	B	32335	B	40922	B	46239	B
	C	53328	B	83834	B	42740	B	53537	B	55200	B
	D	15897	B	21669	B	73244	B	62477	B	62094	B
3	A	78293	B	85150	B	83186	B	94967	B	90712	B
	B	82010	B	76869	B	57154	B	39537	B	22832	D
	C	60064	B	64735	B	65851	B	66712	B	66989	B
	D	79598	B	73028	B	80279	B	59580	B	81218	B
	E	76560	B	84387	B	88394	B	95789	B	96209	B
	F	67008	B	72485	B	73145	B	69675	B	55791	B
	G							19684	B	40587	B
4	A	54004	B	81034	B	74503	B	92375	B	81419	B
	B	1824	B	20608	B	40970	C	15059	B	10960	B
	C	49042	B	52967	B	43902	B	25916	B	47567	B
	D	77385	B	49961	B	67887	B	38701	B	63445	B
	E	86209	B	69002	B	51056	B	80814	B	85249	B
	F	38991	B	36643	B	48124	B	69957	B	48642	B

Table 5c: State level aggregate results for Mathematics for Students with Disabilities.

Claim	Target	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7	
		Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code	Valid N	PL Code
1	A	11811	B	10959	B	5787	B	11982	B	11487	B
	B	5017	A	9220	C	504	B	7121	B	11212	B
	C	11074	B	1826	B	11605	B	11641	B	11385	B
	D	11229	B	12065	B	11028	B	11982	A	11368	B
	E	11518	B	5533	B	12360	B	8924	B	11109	B
	F	8993	B	10175	B	12466	B	11978	B	8052	B
	G	8689	B	12065	B	8608	C	11520	B	8888	B
	H	11811	B	12065	B	5504	B	6745	B	3165	C
	I	4007	B	11626	B	10699	B	3710	B	6646	B
	J	4792	C	1019	B	12427	B	8270	B		
	K	5793	A	7275	A	4584	B				
	L			12054	B						
2	A	11811	B	12065	B	12466	B	11982	B	11510	B
	B	9709	B	837	B	5304	C	6267	B	5128	B
	C	6164	B	11749	B	6931	B	3941	B	7631	B
	D	2499	B	1510	B	8493	B	9411	B	6826	B
3	A	10208	B	11600	B	11125	B	11691	B	10778	B
	B	10851	B	10008	B	5685	B	3135	B	800	D
	C	9711	B	8914	B	10188	B	8542	B	8157	B
	D	10894	B	9720	B	11312	B	8466	A	9845	B
	E	10495	B	11524	B	12355	B	11979	B	11510	B
	F	7619	B	9662	B	8418	B	8497	B	6611	B
	G							2688	B	4819	B
4	A	7588	B	11184	B	9856	B	11357	B	9328	B
	B	320	A	1098	B	4941	C	1617	B	1893	B
	C	7202	B	7991	B	7644	B	4119	B	6111	B
	D	10093	B	6387	B	9356	B	5607	B	8252	B
	E	11491	B	10967	B	7525	B	10369	B	9623	B
	F	4609	B	4085	B	4822	B	7863	B	5399	B

Considerations and Cautions

Unlike the performance levels provided at the total test and claim levels, these strengths and weakness do not imply proficiency. Instead, they show how a group of students' performance is distributed across the content target relatively to their overall performance. For example, a group of students may have performed very well on a subject, but performed lower on a target. Thus, a target performance code of C a target does not necessarily imply a lack of proficiency, but that these students' performance on that target was lower than their performance across other targets put together. In other words, the students performed lower than expected on that target. Although the students are doing well, the educators may still want to focus instruction on the targets with performance code C.

Assessment target score report should serve as a starting point in an overall investigation of students' strengths and weaknesses and constitutes only one of many sources of evidence that should be used in evaluating student performance.

Appendix D: M-STEP SGP and AGP Report

Psychometric Analysis Report for the Michigan English Language Arts (ELA), Mathematics, Science, and SAT Student Growth Percentile and Adequate Growth Percentile Reporting 2018-2019

January 2020

DRC Psychometric Services Michigan Project Team
Data Recognition Corporation

Table of Contents

List of Tables

List of Figures

Introduction

The use of student growth models is common in K-12 testing. The most commonly used approaches by states are conditional growth percentile models, which include student growth percentiles (SGPs, Betebenner, 2008; 2009; 2011) or an alternative known as percentile rank residuals (Castellano & Ho, 2013). Both models attempt to describe individual student growth relative to other students who are academically similar by using prior test scores as predictors. Adequate growth percentiles (AGPs, Betebenner, 2008; 2009; 2011) which use quantile regression models, provide the likelihood students are on track to reaching or maintaining proficiency at some time point in the future. Individual level results from these models can be aggregated at a group level.

SGP analyses were conducted for the M-STEP, SAT, and WIDA, and PRR analysis was conducted for MI-Access assessments. AGP analyses were conducted for M-STEP.

Methodology

Student Growth Percentiles (SGP)

For assessments with a sufficient sample size (M-STEP, SAT, and WIDA Access) student growth percentiles (SGPs) were calculated using the R SGP package (Betebenner et. al., 2015) version 1.9-3.13 as compiled from the master branch of the SGP GitHub repository. SGPs defined this way take a normative approach.

Specially, let Y_t denote an assessment score at time t, the expected value of Y_t at the τ-th quantile, $Q_{Y_t}(\tau|Y_{t-1}, \ldots, Y_1)$ based on prior assessment scores $(Y_{t-1}, \ldots, Y_1)$, is then given by (Betebenner, 2011, p17)

$$Q_{Y_t}(\tau|Y_{t-1}, \ldots, Y_1) = \sum_{j=1}^{t-1}\sum_{i=1}^{3}\phi_{ij}(Y_j)\beta_{ij}(\tau) \quad (1)$$

Where ϕ_{ij}, $i = 1, 2, 3$ and j = 1, …, t-1 denote the B-spline basis functions for quantile τ. For instance, for $\tau = 0.5$, Q_{Y_t} returns the estimated median expectation of Y_t for any combination of $(Y_{t-1}, \ldots, Y_1)$. This analysis used the default parameters of the SGP package which generates 1+7*(number of pretest) parameters per quantile. For example, for a 3-pretest model we have 1+7*3 = 22 parameters per quantile and we estimate 100 quantiles independently (from 0.005 to 0.995 in 0.01 increments).

Calculating a SGP from equation 1 requires prior test score information to determine predicted scores. The SGP for a student is defined as the midpoint of the (ranked) two quantiles between which the student's score falls.

$$SGP_i = \left(\max\left\{\tau_i, \hat{Q}_\tau(Y|X = x_i) < y_i\right\} + \min\left(\left\{\tau_i, \hat{Q}_\tau(Y|X = x_i) > y_i\right\}\right) * \frac{100}{2} \quad (2)$$

Where x_i is the student *i's* vector of prior test scores.

Adequate Growth Percentiles (AGP)

Using the same methodology as described above for calculating SGPs, to calculate a projection or the trajectory a student needs to meet a certain target. An adequate growth percentile, AGP, is the SGP

that a student needs to have to meet or exceed the proficient cut score (or any pre-determined achievement target) within a specified time frame (number of academic years).

Betebenner (2011) contextualizes AGPs in terms of "catch-up", "keep-up", or "move-up." Suppose that an AGP is calculated for a given students Y years away. The following would apply:

> **Catch-Up** is used for students currently not proficient who are expected to reach proficient within *Y* years or by the time they have finished their education, whichever comes first
>
> **Keep-Up** is used for students currently at or above proficient who are expected to remain at or above proficient for all *Y* years or by the time they have finished their education, whichever comes first.
>
> **Move-Up** is used for students currently proficient who are expected to advance beyond proficient within *Y* years or by the time they have finished their education, whichever comes first.

Additionally, a lagged AGP target is also calculated and this value is similar to the AGP. But in this case the current year AGP (i.e. 2019) using the quantile regression model. This gives information to determine if students are on track to reaching proficiency or if they will maintain proficiency over a specified number of years.

Percentile Rank Residuals (PRR)

For assessments with small sample sizes (MI-Access), the PRR method (Castellano & Ho, 2013) was used to estimate the conditional student growth percentiles. This method uses an ordinary least squares (OLS) model, where the predictors consist of past student achievement data.

$$Y_{it} = \beta_0 + \beta_1 y_{i(t-1)} + \beta_2 y_{i(t-2)} + \varepsilon_{it} \quad (5)$$

where Y_{it} is the observed score on the assessment at time *t* for student *i*, $Y_{i,\,t-1}$ is the observed score at prior time 1 and $Y_{i,\,t-2}$ is the observed score at prior time 2.The βs are the regression coefficients, and $\varepsilon_{it}$is a residual error.

After estimating Equation 5, the residuals are calculated using Equation 6:

$$\hat{\varepsilon}_{it} = y_{it} - \hat{y}_{it} \quad (6)$$

where $\hat{\varepsilon}_{it}$is the residual for student *i* at time *t,* $\hat{y}_{it}$is the predicted score from equation 5.

Next, the residuals are rank ordered (Castellano & Ho, 2013, p. 195).

$$PRR_{it} = F(\hat{\varepsilon}_{it}) \times 100 = \frac{\#residuals \leq \hat{\varepsilon}_{it}}{n} \times 100 \quad (7)$$

where $\hat{\varepsilon}_{it}$is the residual for student *i* at time *t* and *n* is the total sample size for all students with MI-Access FI results for a given posttest in 2018-19.

A standard error of measurement can be obtained by simulation for this method. Specifically, for a given posttest, y_{it}, and $CSEM(y_{it})$ 100 posttest were simulated such that they follow a normal distribution given by Equation 8:

$$y_{its} \sim N(mean = y_{it}, sd = CSEM(y_{it})) \quad (8)$$

For each simulated y_{its}, calculate the corresponding PRR using equations 5-7 while holding all other student data constant. Repeat this for each student.

Reporting Results

Results were reported at both the student and aggregate levels. This section provides a brief overview of the results provided to MDE.

For each assessment, results were reported for different content areas. Table 1 provides a list of the grade, assessment, and content area/domain combinations for which SGPs or PRRs were provided.

Table 1: Applicable assessments by grade

Grade	M-STEP	SAT	PSAT	MI-Access	WIDA
K					
1					Overall Composite
2					Overall Composite
3					Overall Composite
4	ELA, Math			ELA, Math	Overall Composite
5	ELA, Math			ELA, Math	Overall Composite
6	ELA, Math			ELA, Math	Overall Composite
7	ELA, Math			ELA, Math, Science	Overall Composite
8	Social Studies		ELA, Math	ELA, Math, Social Studies	Overall Composite
11	Social Studies	ELA, Math		ELA, Math, Science, Social Studies	Overall Composite
12					Overall Composite

AGP Projections

For ELA and Math grades 4 through 7, AGP targets and/or lagged targets were computed for 1 to 3 years from 2019 or 7th grade, whichever comes first. For example, a grade 4 student had AGPs to grades 5, 6, and 7. While a grade 6 student had an AGP to 7th grade. Lagged AGP targets are calculated for Grades 4 through 7. Tables 2 and 3 show the grade progressions for AGP and AGP lagged targets respectively.

Table 2: M-STEP Math and ELA AGP targets by grade, projection year, and grade projected to

Grade	1 Year 2020	2 Year 2021	3 Year 2022
4	5th grade	6th grade	7th grade
5	6th grade	7th grade	
6	7th grade		

Table 3: M-STEP Math and ELA AGP lagged targets by grade and projection year

	Projected AGP Lagged Target Year			
Grade 2018	**Current Year 2019**	**Current +1 Year 2020**	**Current +2 Year 2021**	**Current +3 Year 2022**
3	4th grade	5th grade	6th grade	7th grade
4	5th grade	6th grade	7th grade	
5	6th grade	7th grade		
6	7th grade			

Categorization of Individual (Level) Growth Percentiles

Individual (level) growth percentiles (either SGP or PRR) will also be assigned one of three categorical descriptors based on MDE reporting policies, which are defined as:

- Low: SGP 1-29
- Medium: SGP 30-69
- High: SGP 70-99

Additionally, individual (level) growth percentiles (either SGP or PRR) will also be assigned one of five categorical descriptors based on historical MDE accountability policies. These five categorical descriptors are no longer used in MDE accountability processes but were still calculated for analysis purposes. The five categorical descriptors are defined as:

- Significant Decline (SGP 0-19)
- Decline (SGP 20-39)
- Maintain (SGP 40-59)
- Improvement (SGP 60-79)
- Significant Improvement (SGP 80-99)

Inclusion and Exclusion Rules

Valid Test Sequence Rules

Identified suitable pathways and their information can be found in Table 4 for the SGP method (M-STEP/SAT), the PRR approach (MI-Access FI), and the SGP method (WIDA Access).

Table 4: M-STEP Testing Program Valid Sequence for SGP/AGP calculations

Program	Grade 2019	Prior Year 1	Prior Year 2
M-STEP ELA & Math	4	M-STEP 3rd grade Spring 2018	
	5	M-STEP 4th grade Spring 2018	M-STEP 3rd grade Spring 2017
	6	M-STEP 5th grade Spring 2018	M-STEP 4th grade Spring 2017
	7	M-STEP 6th grade Spring 2018	M-STEP 5th grade Spring 2017
PSAT 8	8	M-STEP 7th grade Spring 2018	M-STEP 6th grade Spring 2017
SAT	11	M-STEP 8th grade Spring 2016	M-STEP 7th grade Spring 2015
M-STEP Social Studies	8	M-STEP 6th grade Spring 2016	
	11	M-STEP 8th grade Spring 2016	M-STEP 6th grade Spring 2014
MI-Access ELA & Math	4	MI-Access 3rd grade Spring 2018	
	5	MI-Access 4th grade Spring 2018	MI-Access 3rd grade Spring 2017
	6	MI-Access 5th grade Spring 2018	MI-Access 4th grade Spring 2017
	7	MI-Access 6th grade Spring 2018	MI-Access 5th grade Spring 2017
	8	MI-Access 7th grade Spring 2018	MI-Access 6th grade Spring 2017
	11	MI-Access 8th grade Spring 2016	MI-Access 7th grade Spring 2015
MI-Access Science	7	MI-Access 4th grade Spring 2016	
	11	MI-Access 7th grade Spring 2015	MI-Access 5th grade Fall 2012
MI-Access Social Studies	8	MI-Access 5th grade Spring 2016	
	11	MI-Access 8th grade Spring 2016	
WIDA	1	WIDA Kindergarten Spring 2018	
	2	WIDA 1st grade Spring 2018	WIDA Kindergarten Spring 2017
	3	WIDA 2nd grade Spring 2018	WIDA 1st grade Spring 2017
	4	WIDA 3rd grade Spring 2018	WIDA 2nd grade Spring 2017
	5	WIDA 4th grade Spring 2018	WIDA 3rd grade Spring 2017
	6	WIDA 5th grade Spring 2018	WIDA 4th grade Spring 2017
	7	WIDA 6th grade Spring 2018	WIDA 5th grade Spring 2017
	8	WIDA 7th grade Spring 2018	WIDA 6th grade Spring 2017
	9	WIDA 8th grade Spring 2018	WIDA 7th grade Spring 2017
	10	WIDA 9th grade Spring 2018	WIDA 8th grade Spring 2017
	11	WIDA 10th grade Spring 2018	WIDA 9th grade Spring 2017
	12	WIDA 11th grade Spring 2018	WIDA 10th grade Spring 2017

Minimum Number of Students

A minimum of 5,000 students were required for the SGP M-STEP & SAT run.

A minimum of 1,000 students was preferred for the MI-Access FI PRR run.

A minimum of 2,000 students were required for the SGP WIDA Access for ELLs 2.0 run.

Repeat Test Takers

Students who repeated the grade immediately before the posttest were not included in either the SGP or the PRR analysis, thus the SGPs were not calculated for these students. For instance, if posttest score (Y_t) and prior 1 year score (Y_{t-1}) are with the same grade, the student was not included in the analysis and does not receive an SGP.

Skipped Grades

Students who skipped the grade immediately prior to the posttest were not included in the analysis (i.e. 5th grade posttest following skipping 4th grade in the previous example.) In addition, if a student has a test sequence with a skipped grade, only the grade prior will be used to calculate the SGP.

Gaps in Test Sequence

Some students in the dataset are missing certain years of test scores. This may be due to student mobility, missed test windows, or other factors (e.g., Grade 3 M-STEP ELA in Spring 2017, followed by Grade 5 M-STEP ELA in Spring 2019). Students with a gap were not included unless they have a recent, valid sequence leading up to the posttest.

Home School and Private School Exclusion

All home schooled and private school test records were excluded from computing SGP. MDE will ensure that students who were previously tested as home schooled or at a private school are also excluded from the data pull.

Student Level Results for SGPs and PRRs

Student level results provided to MDE for SGPs and PRRs included:

1. Demographic and assessment information
2. SGPs
3. SGP standard errors
4. SGP Growth Level Code
5. SGP Norm Group
6. Estimation Method
7. Prior achievement information used

Student Level Results for AGPs

Student level results provided to MDE for AGPs included:

1. Demographic and assessment information
2. AGP Years Projected (1-4)
3. AGP Target
4. AGP Lagged Target
5. AGP Stay/Move Up Target
6. AGP Lagged Stay/Move Up Target

Aggregation

Results were aggregated by assessment and accountability at the state, district, and school level using a variety of subgroups specified by MDE. Aggregation results included:

1. Count of students included
2. Average (arithmetic mean) of the SGPs
3. Standard deviation of SGPs
4. Count of students at each of five growth levels (Significant Improvement, Improvement, Maintain, Decline, Significant Decline)
5. Percentage of students at each of these five levels as a percentage of total students with SGPs
6. Count of students at each of three growth levels (Low, Medium, High)
7. Percentage of students at each of these three levels as a percentage of total students with SGPs.
8. Building z-score

Quality Control

DRC's psychometric team verified the data coming from MDE followed the rules, structure, and specifications agreed upon by both DRC and MDE. Any issues around unexpected data or missing fields were addressed by MDE.

To ensure that the proper growth model was used, base R code was written by the psychometrician and verified by a consultant and a statistical analyst. The code for each subject was reviewed and SGP, PRR, or AGP values were internally checked for reasonability. Two staff members from the psychometric services team verified aggregate results by independent replication, and MDE reviewed the reasonability of the aggregate and individual SGP, PRR, or AGP results. Results went through several iterations of independent replication and MDE review until all discrepancies were resolved.

Summary of Results

Tables 5 through 9 provide a summary of the number of students and median growth SGPs or PRR values by aggregate levels. Tables 5 and 6 provide the summary of number of students and median growth (SGP or PRR) by testing program, calculation method, content area, and grade. Table 7 provides the results by calculation method, content area, and grade. Table 8 provides the results by content area and grade and Table 9 provides the results by grade. As expected with these methods, the median values tend to be near 50.

Table 5: Number of cases and median SGP by testing program, content area, and grade.

Testing Program	Content Area	Grade	N	Median
M-STEP	English Language Arts	4	98,372	50
		5	101,393	50
		6	104,787	50
		7	98,372	50
	Mathematics	4	98,609	50
		5	101,570	50
		6	104,942	50
		7	104,751	50
	Social Studies	8	98,160	49
		11	91,542	50
PSAT	English Language Arts	8	103,386	50
	Mathematics	8	103,488	50
SAT	English Language Arts	11	91,751	50
	Mathematics	11	91,829	50
WIDA	WIDA	1	7,639	50
		2	8,829	50
		3	8,877	50
		4	8,326	51
		5	7,173	51
		6	5,884	51
		7	5,978	51
		8	5,627	50
		9	5,351	51
		10	4,924	50
		11	4,287	50
		12	3,300	50

Table 6: Number of cases and median PRR by testing program, content area, and grade.

Testing Program	Content Area	Grade	N	Median
MI-Access	English Language Arts	4	835	50
		5	929	51
		6	1,095	51
		7	1,108	51
		8	1,129	51
		11	872	50
	Mathematics	4	872	50
		5	1,010	51
		6	1,195	51
		7	1,238	50
		8	1,250	50
		11	953	51
	Science	7	854	50
		11	889	49
	Social Studies	8	956	50
		11	954	50

Table 7: Number of cases and median growth by method, content area, and grade.

Method	Content Area	Grade	N	Median
PRR	English Language Arts	4	835	50
		5	929	51
		6	1,095	51
		7	1,108	51
		8	1,129	51
		11	872	50
	Mathematics	4	872	50
		5	1,010	51
		6	1,195	51
		7	1,238	50
		8	1,250	50
		11	953	51
	Science	7	854	50
		11	889	49
	Social Studies	8	956	50
		11	954	50
SGP	English Language Arts	4	98,372	50
		5	101,393	50
		6	104,787	50
		7	104,635	50
		8	103,386	50
		11	91,751	50
	Mathematics	4	98,609	50
		5	101,570	50
		6	104,942	50
		7	104,751	50
		8	103,488	50
		11	91,829	50
	Social Studies	8	98,160	49
		11	91,542	50

Table 8: Number of cases and median growth by content area and grade.

Content Area	Grade	N	Median
English Language Arts	4	99,207	50
	5	102,322	50
	6	105,882	50
	7	105,743	50
	8	104,515	50
	11	92,623	50
Mathematics	4	99,481	50
	5	102,580	50
	6	106,137	50
	7	105,989	50
	8	104,738	50
	11	92,782	50
Science	7	854	50
	11	889	49
Social Studies	8	99,116	49
	11	92,496	50

Table 9: Number of cases and median growth by grade.

Grade	N	Median
1	7,639	50
2	8,829	50
3	8,877	50
4	207,014	50
5	212,075	50
6	217,903	50
7	218,564	50
8	313,996	50
9	5,351	51
10	4,924	50
11	283,077	50
12	3,300	50

Goodness of Fit

To examine the fit of the growth models, the correlations between the outcome score (2018) and the prior achievement score was calculated. Tables 10 and 11 provide the correlations by program, content area, and grade. All correlations are acceptable and within the moderate range. For the M-STEP program, all correlations are consistent within content area. In Mathematics and English Language Arts, correlations are at or above 0.80, for Social Studies it is at or above 0.75. With the SAT correlations similar with a correlation of 0.81 for English Language Arts and Mathematics. WIDA correlations are fairly consistent but lower, ranging from 0.62 to 0.80. Finally, the correlations for MI-Access are

consistent within content area but lower ranging from 0.50 to 0.67 for English Language Arts, from 0.50 to 0.61 for Mathematics, 0.46 to 0.53 for Science and 0.46 to 0.53 for Social Studies.

Table 10: Correlation between current SS and prior SS by testing program, content area, and grade for SGP models.

Testing Program	Content Area	Grade	N	Correlation
M-STEP	English Language Arts	4	98,372	0.82
		5	101,393	0.84
		6	104,787	0.84
		7	104,635	0.85
	Mathematics	4	98,609	0.85
		5	101,570	0.86
		6	104,942	0.86
		7	104,751	0.88
	Social Studies	8	98,160	0.75
		11	91,542	0.76
PSAT	English Language Arts	8	103,386	0.80
	Mathematics	8	103,488	0.84
SAT	English Language Arts	11	91,751	0.81
	Mathematics	11	91,829	0.81
WIDA	WIDA	1	7,639	0.62
		2	8,829	0.74
		3	8,877	0.78
		4	8,326	0.75
		5	7,173	0.76
		6	5,884	0.73
		7	5,978	0.78
		8	5,627	0.80
		9	5,351	0.75
		10	4,924	0.77
		11	4,287	0.75
		12	3,300	0.67

Table 11: Correlation between current SS and prior SS by testing program, content area, and grade for PRR model.

Testing Program	Content Area	Grade	N	Correlation
MI-Access	English Language Arts	4	835	0.56
		5	929	0.62
		6	1,095	0.65
		7	1,108	0.67
		8	1,129	0.65
		11	872	0.56
	Mathematics	4	872	0.50
		5	1,010	0.61
		6	1,195	0.56
		7	1,238	0.58
		8	1,250	0.60
		11	953	0.55
	Science	7	854	0.46
		11	889	0.53
	Social Studies	8	956	0.46
		11	954	0.53

Distributions of SGPs and PRRs

The distributions of SGPs and PRRs are provided in Figure 1 through Figure 3, which shows that SGPs tend to uniformly range from 1 to 99. While the PRRs also range from 1 to 99, they are a bit less stable due to the small sample sizes used in the calculations. It should be noted that the differences in the distributions of PRRs and SGPs across grade and content area tend to be relatively small given the scale of the density plots range from 0 to 0.012.

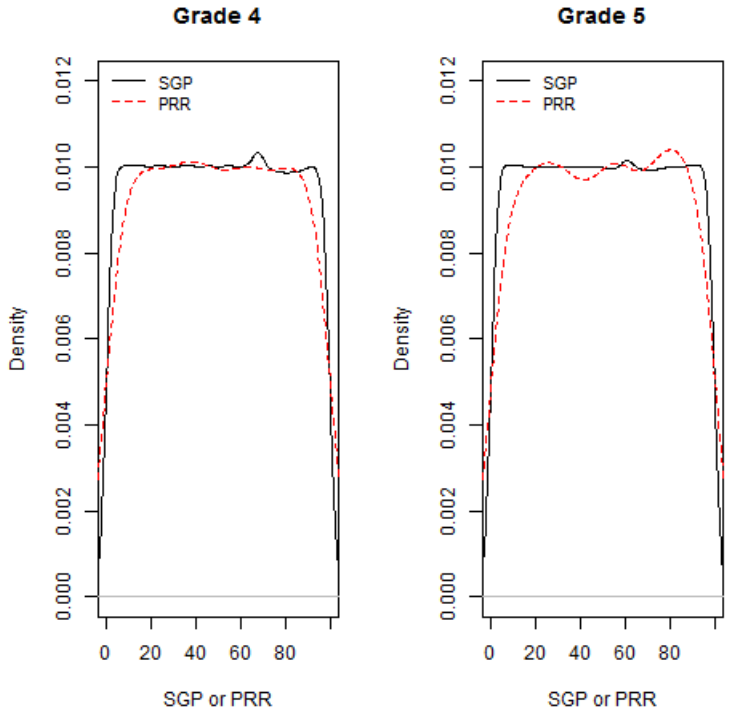


Figure 1. Distribution of SGP/PRR for Mathematics Grades, 4 and 5

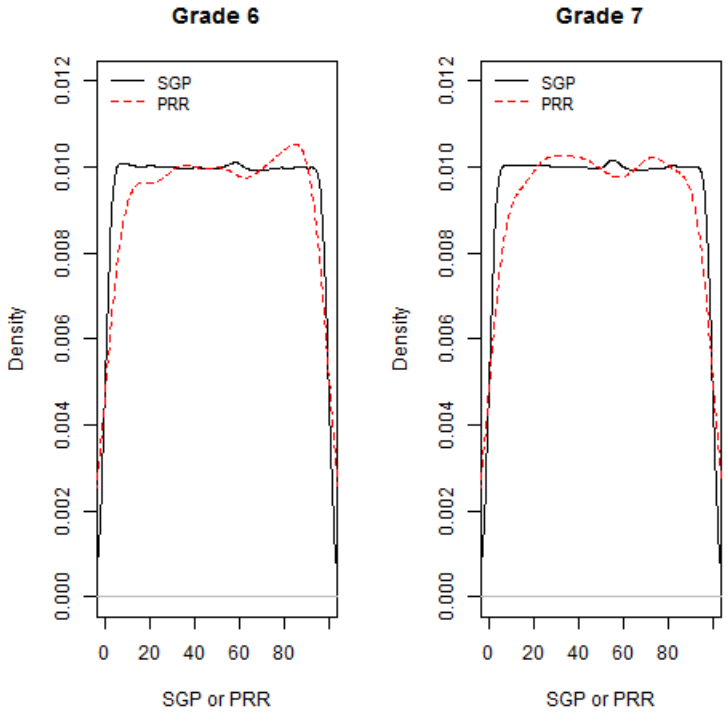


Figure 2. Distribution of SGP/PRR for Mathematics Grades, 6 and 7

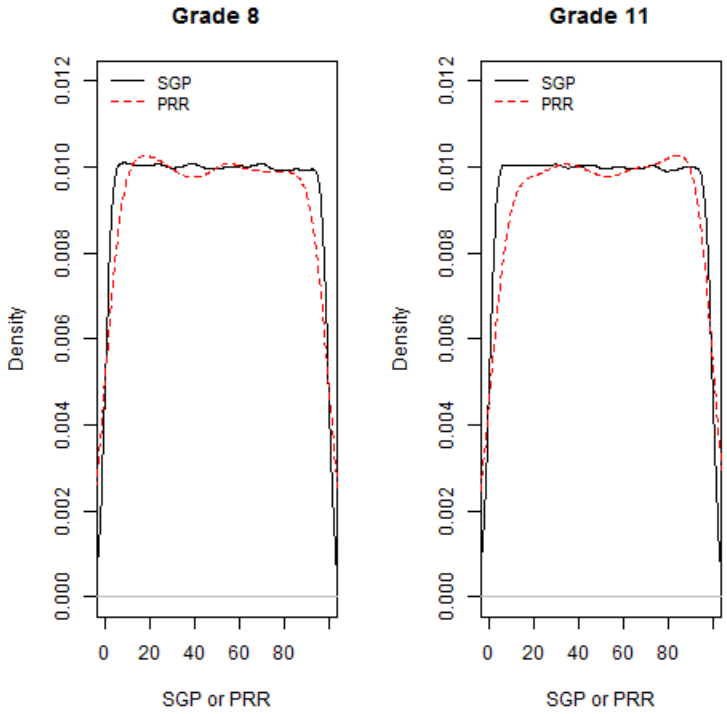


Figure 3. Distribution of SGP/PRR for Mathematics Grades, 8 and 11

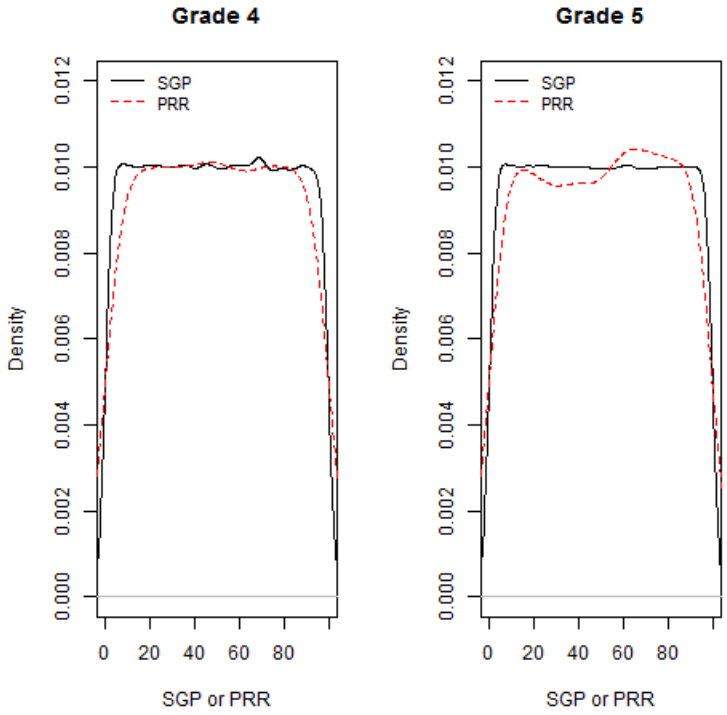


Figure 4. Distribution of SGP/PRR for English Language Arts Grades, 4 and 5

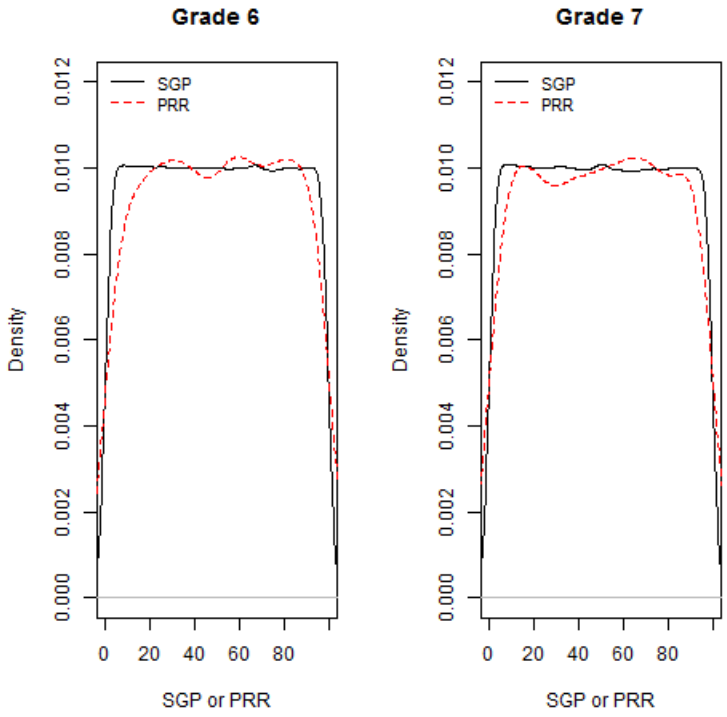


Figure 5. Distribution of SGP/PRR for English Language Arts Grades, 6 and 7

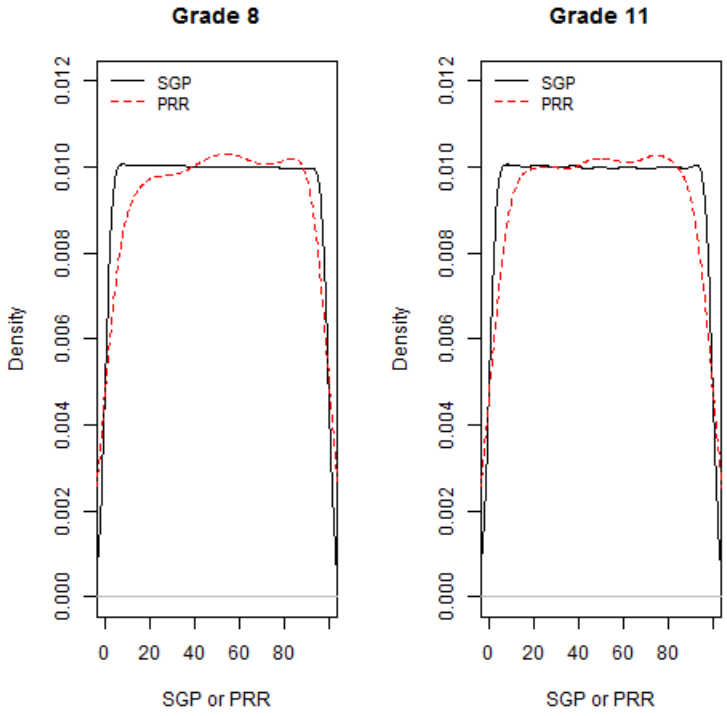


Figure 6. Distribution of SGP/PRR for English Language Arts Grades, 8 and 11

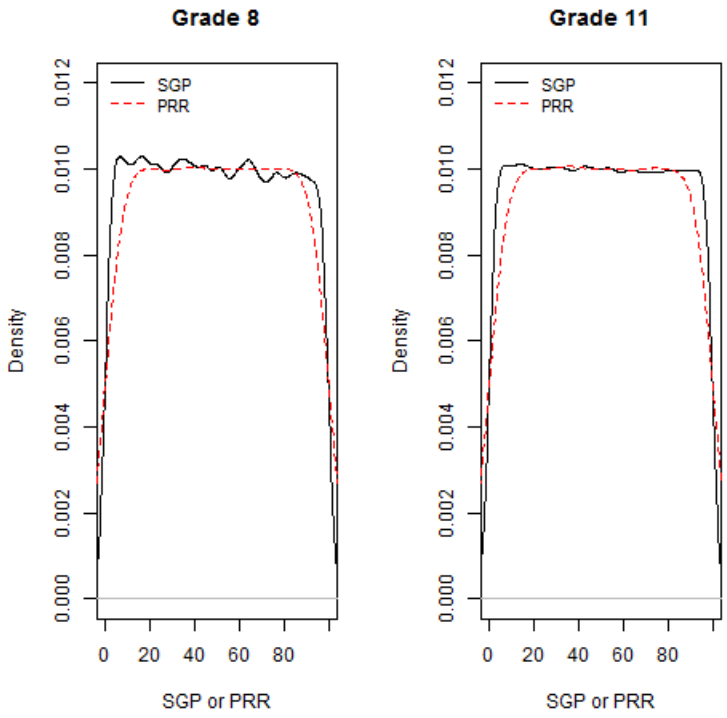


Figure 7. Distribution of SGP/PRR for Social Studies Grades, 8 and 11

Checks for Neutrality

Since the growth models used in this analysis do not control for demographic variables, particularly those that may have some impact on student growth rates and trajectories, it is unknown whether the results are biased, especially when aggregated at the school or district level (Education Analytics, 2015). Thus, it is important to look at the relationship between the aggregated growth measure, in this case median SGP and the variables of interest that were not controlled for in the growth models. It is important to note that it is unknown what the correlations "should be." Tables 12 and 13 provide the correlations between the median SGP for a school or a district (with more than 20 students) related to the percentage of each demographic for that building or district. Graphs of these relationships can be found in the appendix.

Table 12: Correlations between Median SGP and Demographic at the school level. [1]

Content Area	ED	SE	EL	Non-White
English Language Arts	-0.39	-0.17	0.04	-0.24
Mathematics	-0.38	-0.16	0.04	-0.23
Science				
Social Studies	-0.41	-0.16	-0.08	-0.25
WIDA	-0.38	-0.02		-0.19

Table 13: Correlations between Median SGP and Demographic at the district level.

Content Area	ED	SE	EL	Non-White
English Language Arts	-0.38	-0.27	0.01	-0.17
Mathematics	-0.37	-0.18	0.01	-0.19
Science	-0.16		-0.57	0.33
Social Studies	-0.37	-0.21	-0.01	-0.19
WIDA	-0.35	-0.02		-0.11

When aggregating growth model outcomes, it is also important to note that growth models, as with most regression models, have issues (more variability or less precision) when sample sizes are small. This is also true when aggregating growth model results at the school level. Figure 8 provides the relationship between the number of students and SGP. This shows that there is less variability in median SGP as the number of students increase.

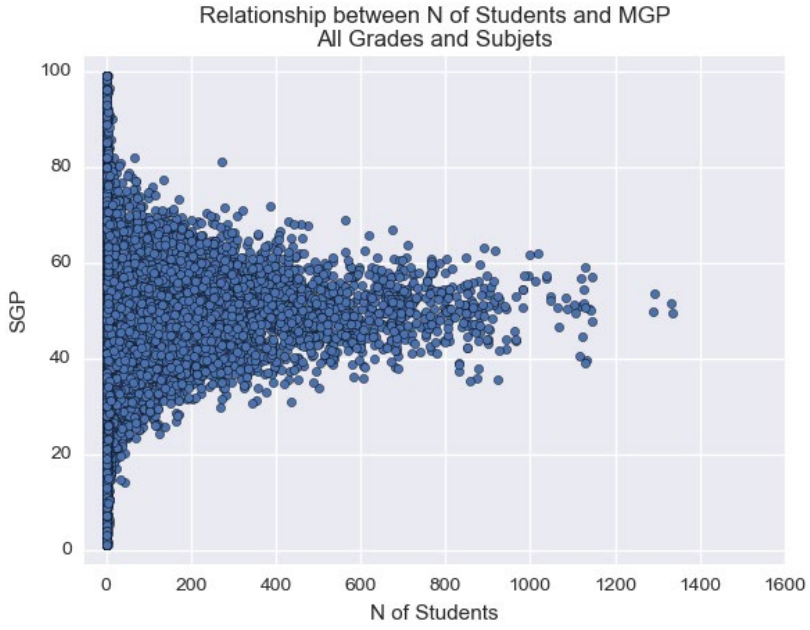


Figure 8. Number of Students versus Median SGP

[1] Since Science was administered only for MI-Access, there were no schools with more than 20 students

AGP Outcomes

In 2019, AGPs and target AGPs were computed for M-STEP ELA and Mathematics, grades 4 through 7. The number of years projected in the model was varied between 1 and 4. Details can be found in Tables 2 and 3. One way to aggregate these results is to compare the percentage of students meeting targets by their 2019 performance level, grade, and years projected. Tables 14 and 15 do this by showing the percentage of students, by grade, who have a 2019 SGP greater than their 2019 lagged AGP, broken down by proficiency level, grade, and years projected. For example, in Grade 4 ELA, 65% of proficient students are on track to remain proficient (or reach advanced) in three years' time. These tables show that students who end in the highest performance level (Advanced) do so because they consistently grew at levels surpassing that which was necessary to achieve and maintain proficiency. Similarly, they also show that students who end in the lowest performance level (Not Proficient) do so because they consistently grew at levels well below what was necessary to reach proficiency.

Table 14: Percentage of students whose 2019 SGP exceeds their lagged by performance level and years projected for M-STEP ELA.

		Not Proficient		Partially Proficient		Proficient		Advanced	
Grade	Years Projected	N Total	% 2018 SGP Exceeds Lagged AGP	N Total	% 2018 SGP Exceeds Lagged AGP	N Total	% 2018 SGP Exceeds Lagged AGP	N Total	% 2018 SGP Exceeds Lagged AGP
4	1	32,318	0%	20,517	16%	21,422	86%	24,115	100%
	2	32,318	0%	20,517	27%	21,422	69%	24,115	97%
	3	32,318	2%	20,517	34%	21,422	65%	24,115	94%
	4	32,318	2%	20,517	34%	21,422	65%	24,115	94%
5	1	32,245	0%	21,957	9%	29,093	81%	18,098	100%
	2	32,245	1%	21,957	23%	29,093	72%	18,098	99%
	3	32,245	1%	21,957	23%	29,093	72%	18,098	99%
	4	32,245	1%	21,957	23%	29,093	72%	18,098	99%
6	1	32,706	0%	27,984	11%	29,782	86%	14,315	100%
	2	32,706	0%	27,984	11%	29,782	86%	14,315	100%
	3	32,706	0%	27,984	11%	29,782	86%	14,315	100%
	4	32,706	0%	27,984	11%	29,782	86%	14,315	100%
7	1	30,397	0%	28,970	0%	32,009	95%	13,259	100%
	2	30,397	0%	28,970	0%	32,009	95%	13,259	100%
	3	30,397	0%	28,970	0%	32,009	95%	13,259	100%
	4	30,397	0%	28,970	0%	32,009	95%	13,259	100%

Table 15: Percentage of students whose 2019 SGP exceeds their lagged by performance level and years projected for M-STEP Math.

		Not Proficient		Partially Proficient		Proficient		Advanced	
Grade	Years Projected	N Total	% 2018 SGP Exceeds Lagged AGP	N Total	% 2018 SGP Exceeds Lagged AGP	N Total	% 2018 SGP Exceeds Lagged AGP	N Total	% 2018 SGP Exceeds Lagged AGP
4	1	23,787	0%	33,164	3%	25,149	72%	16,509	100%
	2	23,787	0%	33,164	12%	25,149	67%	16,509	99%
	3	23,787	0%	33,164	19%	25,149	65%	16,509	97%
	4	23,787	0%	33,164	19%	25,149	65%	16,509	97%
5	1	36,530	0%	29,314	11%	18,440	82%	17,286	100%
	2	36,530	0%	29,314	24%	18,440	75%	17,286	99%
	3	36,530	0%	29,314	24%	18,440	75%	17,286	99%
	4	36,530	0%	29,314	24%	18,440	75%	17,286	99%
6	1	35,401	0%	32,306	10%	20,080	86%	17,155	100%
	2	35,401	0%	32,306	10%	20,080	86%	17,155	100%
	3	35,401	0%	32,306	10%	20,080	86%	17,155	100%
	4	35,401	0%	32,306	10%	20,080	86%	17,155	100%
7	1	36,943	0%	29,872	1%	20,471	94%	17,465	100%
	2	36,943	0%	29,872	1%	20,471	94%	17,465	100%
	3	36,943	0%	29,872	1%	20,471	94%	17,465	100%
	4	36,943	0%	29,872	1%	20,471	94%	17,465	100%

References

Betebenner, D. W. (2011). *A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories.* Downloaded March 9, 2018 from http://www.nj.gov/education/njsmart/performance/SGP_Technical_Overview.pdf.

Betebenner, D. W., VanIwaarden, A., Domingue, B., and Shang, Y. (2016). SGP: An R Package for the Calculation and Visualization of Student Growth Percentiles & Percentile Growth Trajectories. R package version 1.5-0.0. https://cran.r-project.org/web/packages/SGP/

Castellano, K.E., and Ho, A.D. (2015). Practical Differences Among Aggregate-Level Conditional Status Metrics: From Median Student Growth Percentiles to Value-Added Models. Journal of Educational and Behavioral Statistics, 40(1), 35-68. doi:10.3102/107699861454848

Castellano, K.E., and Ho, A.D. (2013). Contrasting OLS and Quantile Regression Approaches to Student "Growth" Percentiles. Journal of Educational and Behavioral Statistics, 38(2), 190-215. doi:10.3102/1076998611435413

Education Analytics (2015). Michigan Department of Education Technical Report.

Appendix

English Language Arts

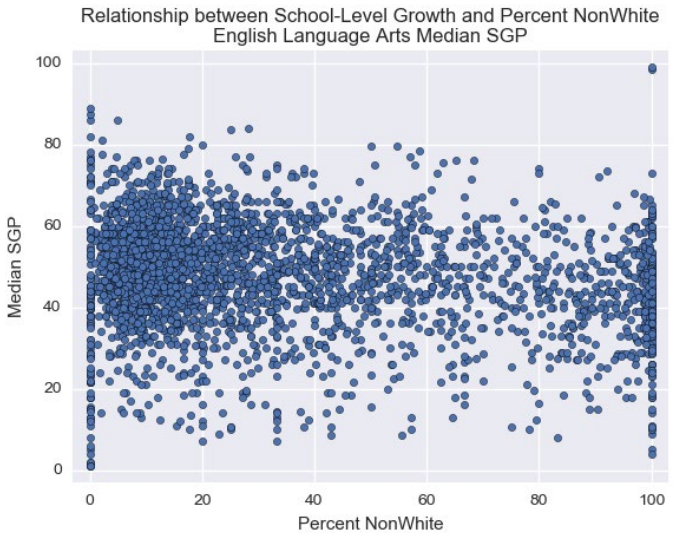


Figure A.1. Median School SGP versus Percentage of Non-White Students for English Language Arts

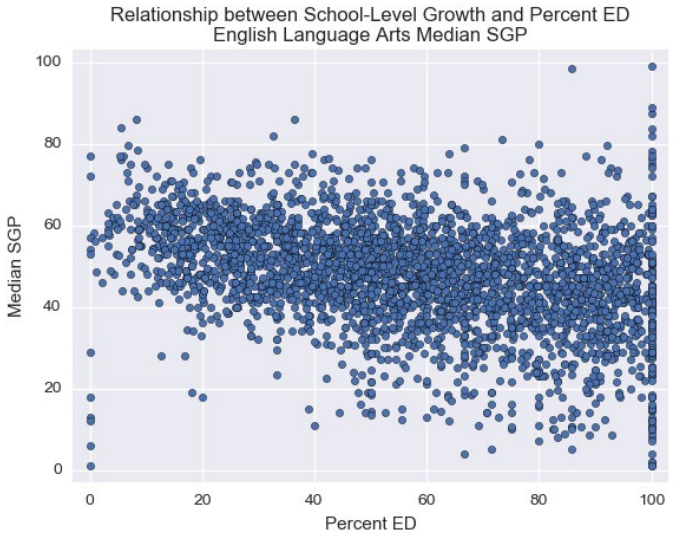


Figure A.2. Median School SGP versus Percentage of Economically Disadvantaged Students for English Language Arts

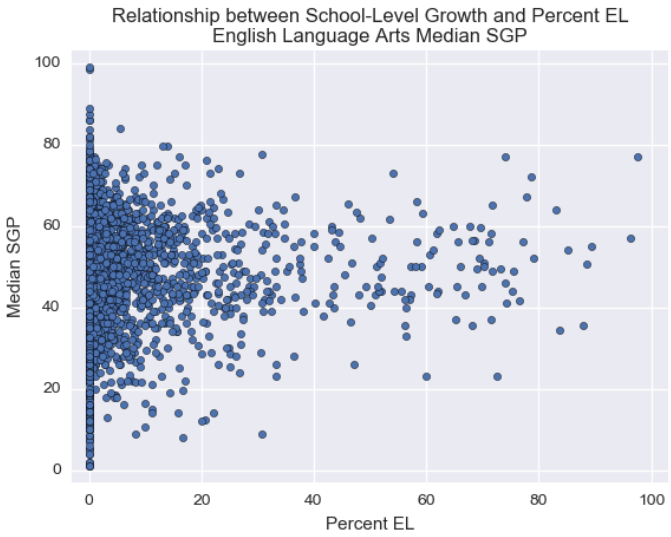


Figure A.3. Median School SGP versus Percentage of English Learner (EL) Students for English Language Arts

Figure A.4. Median School SGP versus Percentage of Special Education Students for English Language Arts

Mathematics

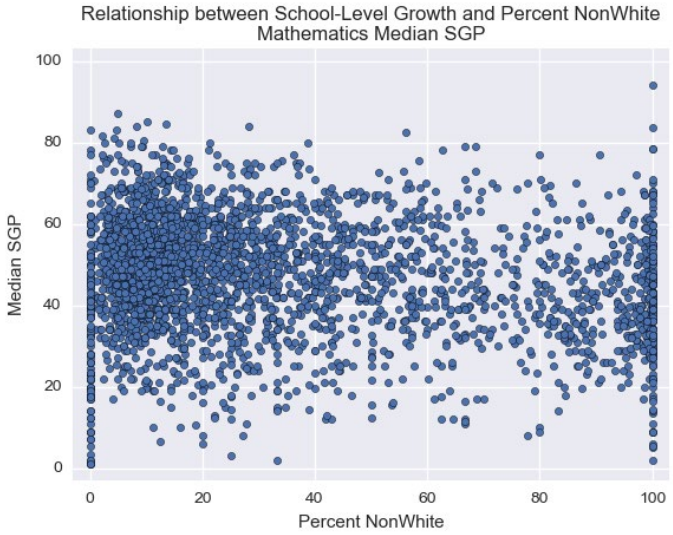


Figure A.5. Median School SGP versus Percentage of Non-White Students for Mathematics

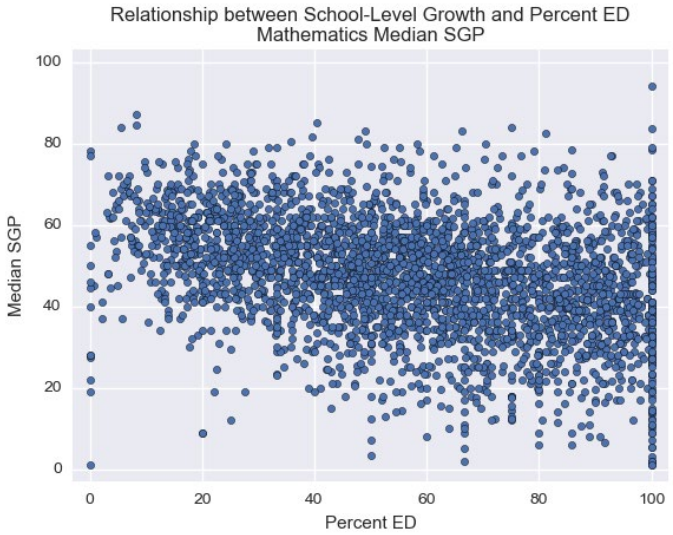


Figure A.6. Median School SGP versus Percentage of Economically Disadvantaged for Mathematics

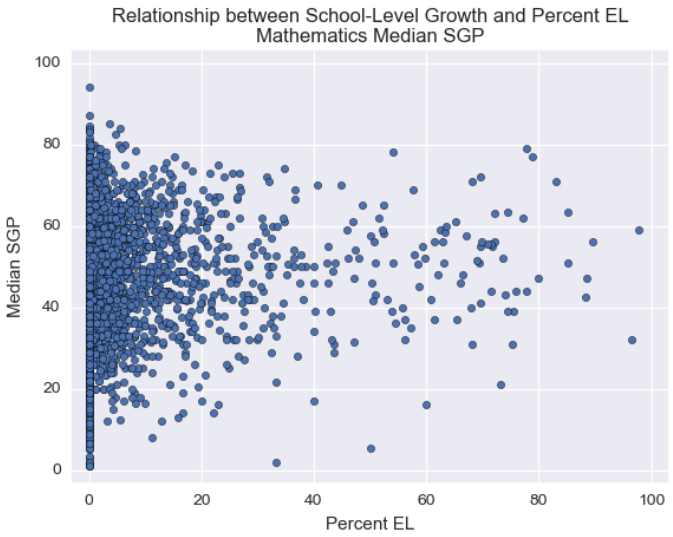


Figure A.7. Median School SGP versus Percentage of English Learner (EL) Students for Mathematics

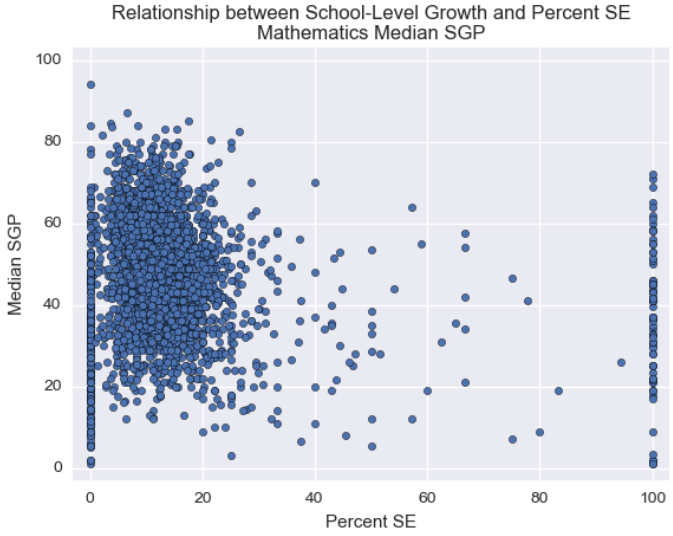


Figure A.8. Median School SGP versus Percentage of Special Education Students for Mathematics

Science

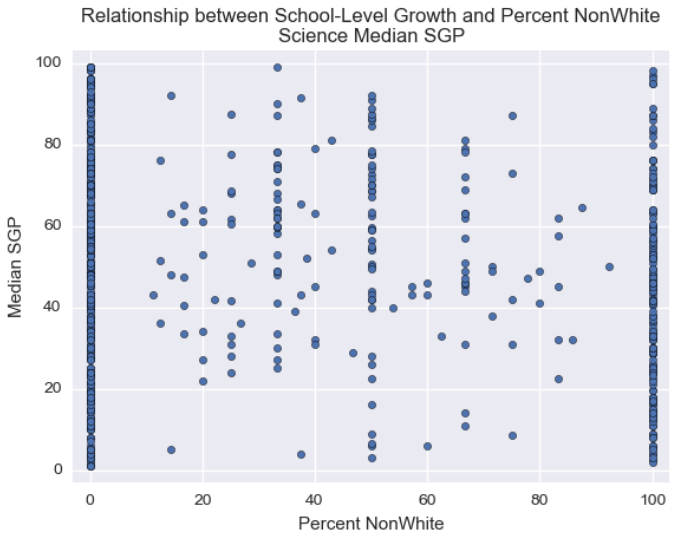


Figure A.9. Median School SGP versus Percentage of Non-White Students for Science[2]

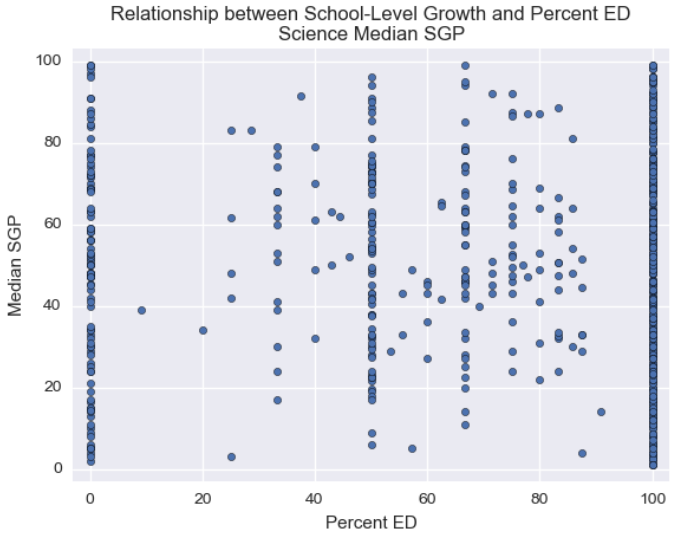


Figure A.10. Median School SGP versus Percentage of Economically Disadvantaged Students for Science[2]

[2] Note that MI-Access is the only science assessment addressed in this report.

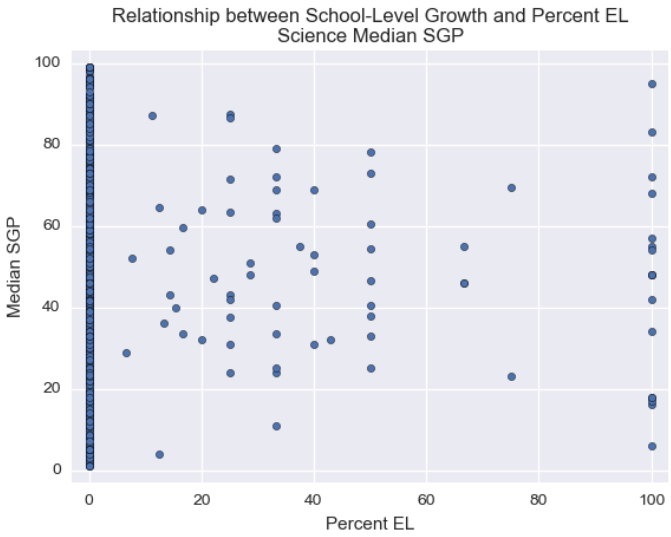


Figure A.11. Median School SGP versus Percentage of English Learner (EL) Students for Science[3]

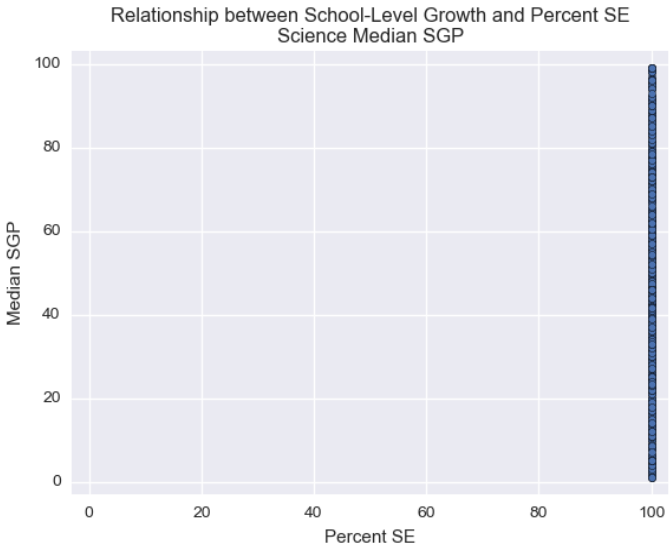


Figure A.12. Median School SGP versus Percentage of Special Education Students for Science[3]

[3] Note that MI-Access is the only science assessment addressed in this report.

Social Studies

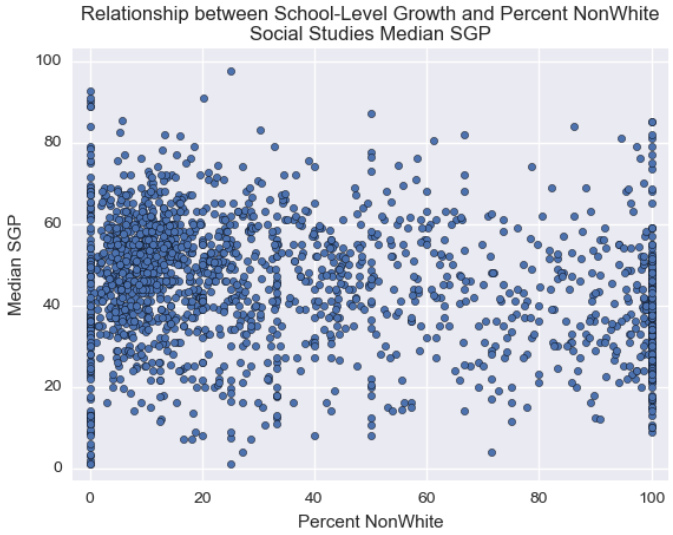


Figure A.13. Median School SGP versus Percentage of Non-White Students for Social Studies

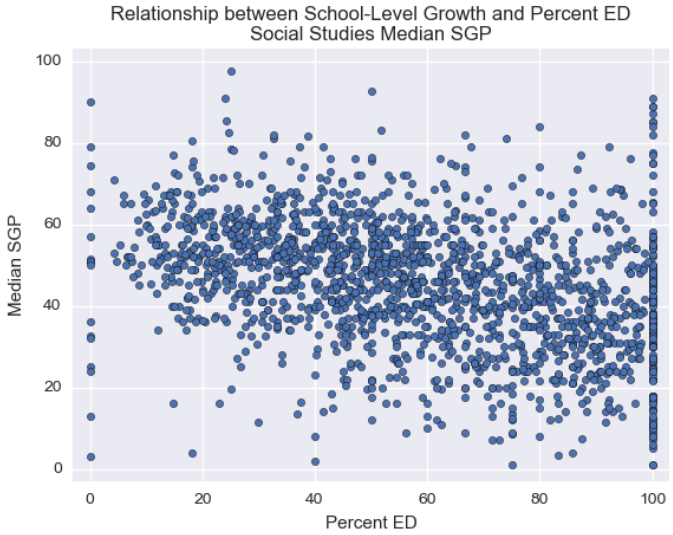


Figure A.14. Median School SGP versus Percentage of Economically Disadvantaged Students for Social Studies

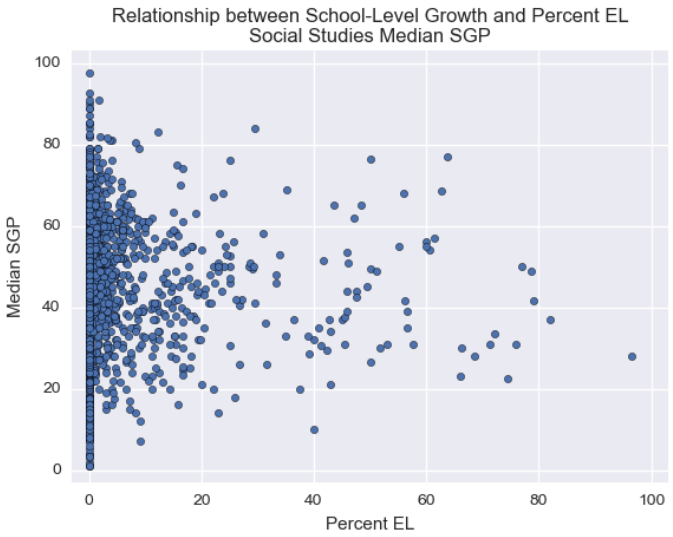


Figure A.15. Median School SGP versus Percentage of English Learner (EL) Students for Social Studies

Figure A.16. Median School SGP versus Percentage of Special Education Students for Social Studies

WIDA

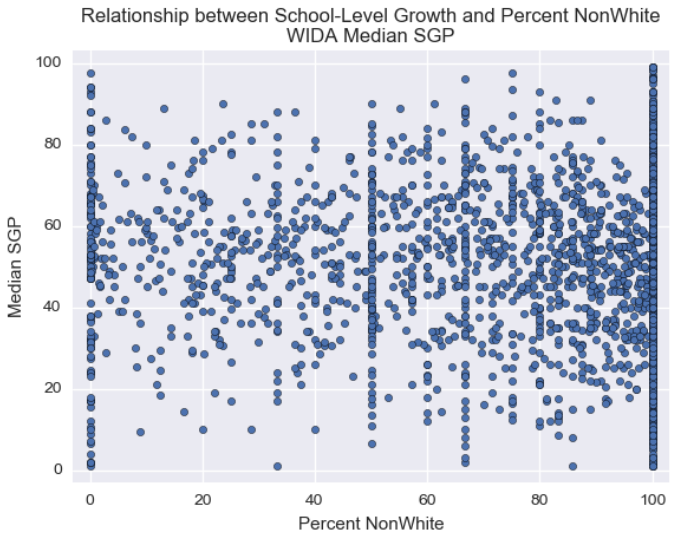


Figure A.17. Median School SGP versus Percentage of Non-White Students for WIDA[4]

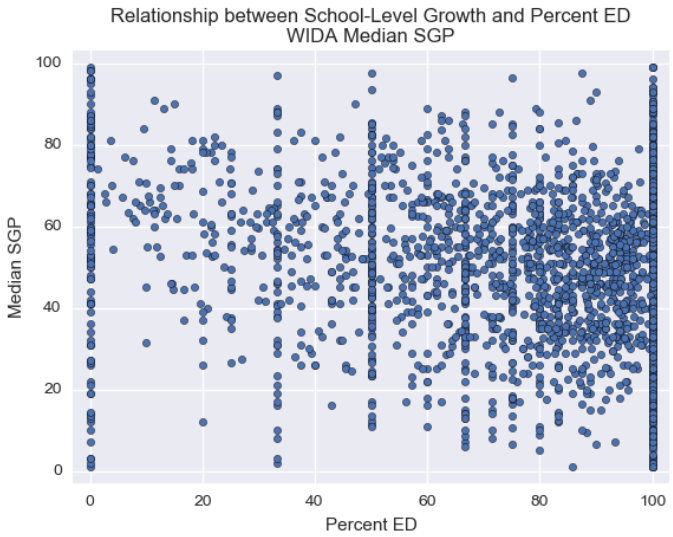


Figure A.18. Median School SGP versus Percentage of Economically Disadvantaged Students for WIDA[4]

[4] Note that the WIDA assessment is administered to only English Learner (EL) students

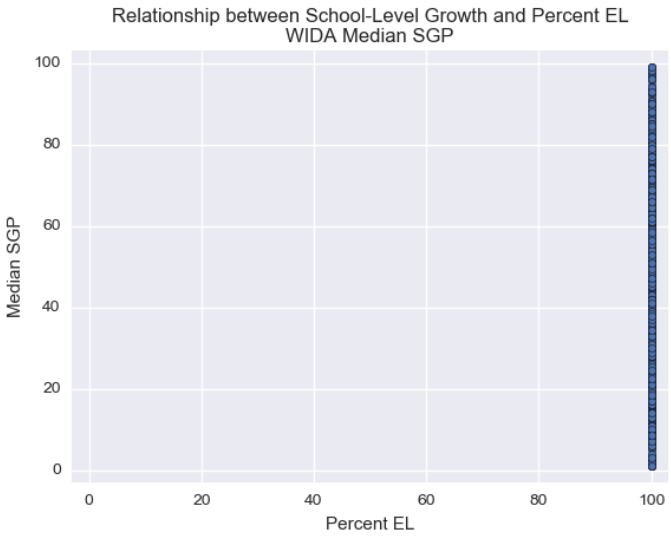


Figure A.19. Median School SGP versus Percentage of English Learner (EL) Students for WIDA[5]

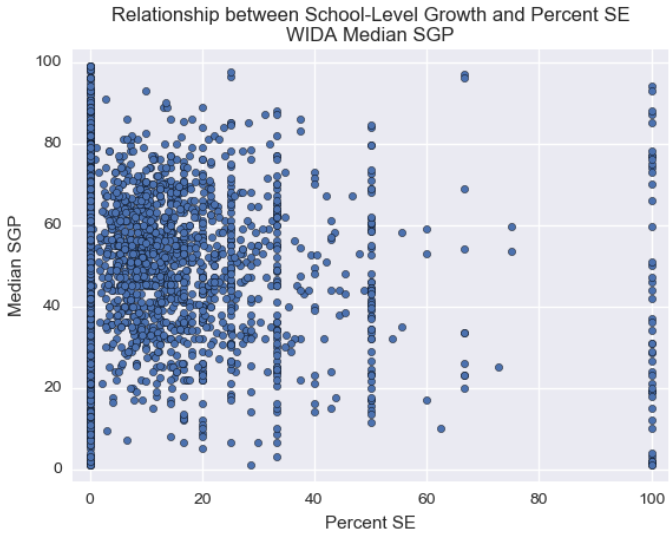


Figure A.20 Median School SGP versus Percentage of Special Education Students for WIDA[5]

[5] Note that the WIDA assessment is administered to only English Learner (EL) students

Appendix E: M-STEP Standards Validation

Appendix E-1. Validity Evidence for English Language Arts and Mathematics Cut Scores

Ricardo Mercado, Jessalyn Smith, Sara Kendallen, Mayuko Simon, Alassane Savadogo, and Ben Sorenson *Data Recognition Corporation*

July 15, 2018

Appendix E-2 Summary

- On July 9–12, 2018, the Michigan Department of Education (MDE) partnered with Data Recognition Corporation (DRC) to conduct a standards validation for the Michigan Student Test of Educational Progress (M-STEP) tests of English language arts (ELA) and mathematics for grades 3–8.
- The *standards validation* was needed because of test-length reductions implemented in spring 2018. Specifically, proportional reductions in the number of items by reporting category were implemented for mathematics; and for ELA, new passage-based writing items replaced other performance tasks.
- The purpose of the standards validation workshop was to determine whether the existing M STEP cut scores were still valid for continued use on the updated tests.
- Participants' recommendations at the standards validation were consistent with the existing cut scores, providing evidence of their validity for continued use.

Appendix E-3 Background

The M-STEP is administered to assess Michigan students' mastery of the Michigan Academic *Standards*. The assessments began as an implementation of the Smarter Balanced Assessment Consortium's (SBAC) ELA and mathematics tests. The current cut scores for the tests are taken from the SBAC tests.

Over the course of several years, important changes have been made to the assessments to make them more meaningful to Michigan educators. These include the alignment of the test items to the Michigan Academic *Standards*, the implementation of a Michigan-specific test blueprint, and a reduction in the number of performance tasks used in ELA to reduce overall test time. These changes were made cautiously and deliberately with the active involvement of Michigan educators and stakeholders.

In school year 2017–18, the tests in grades 3–8 were shortened to reduce the time burden on students and schools. To do so, all performance tasks in ELA were replaced with passage-based writing items, a new item type for Michigan. The ELA test blueprints were adjusted to accommodate the new item type and the reduction in test length. In grades 3–8 mathematics, the test was also shortened to reduce overall testing time, but this change did not involve adding new test items or significantly altering the test blueprint.

Appendix E-4 Standards Validation Methodology

The purpose of the standards validation was to determine whether the current M-STEP cut scores for grades 3–8 ELA and mathematics were still valid for continued use, given the 2018 updates to the tests.

A total of 54 Michigan educators engaged in a modification of the Bookmark Standard Setting Procedure (Lewis, Mitzel, & Green, 1996; Lewis, Mitzel, Mercado, & Schulz, 2012) to validate the cut scores. This method has been used on large-scale assessments in Michigan and across the nation, including for SBAC.

Participants studied the existing Michigan performance level descriptors (PLDs) and Michigan Learning *Standards* to review the knowledge, skills, and abilities expected of students in each performance level. The four performance levels on M-STEP are *Not Proficient*, *Partially Proficient*, *Proficient*, and *Advanced*. Each performance level is associated with a level of mastery of the Michigan Learning *Standards*. Participants then discussed the content-based expectations for students at the threshold of each performance level (e.g., a student who is just *Proficient*). To support their discussions of these threshold students, participants were provided with the SBAC achievement level descriptors (ALDs). These SBAC ALDs were used at the original standard setting where the cut scores were established.

Participants studied collections of test items that were ordered in terms of difficulty. The existing cut scores were presented as benchmarks for participants' consideration: participants were asked to consider the knowledge and skills that students would need to demonstrate on the updated ELA and mathematics tests, as based on the benchmarked (existing) cut scores. Then, participants compared these expectations against the content-based expectations for students at the thresholds of each performance level. Participants were instructed to recommend retaining the existing cut scores if there was good correspondence between the benchmarks and these content-based expectations, or to recommend alternative cut scores that reflect better correspondence. Participants engaged in two rounds of individual judgments and group discussion. (The grade 5 mathematics committee engaged in three rounds of judgments to accommodate additional discussion.) The committees' median judgments were taken as their final recommendations.

Before the workshop, it was hypothesized that participants would recommend cut scores which were similar to, but not exactly equal to, the existing cut scores. The rationale behind this hypothesis was that nearly any group of educators going through an iterative, judgmental process like the Bookmark Procedure will tend to arrive at slightly different judgments at the end of the process. Accordingly, it was not expected that standards validation participants would recommend cut scores exactly equal to the existing cut scores: slight differences in cut score recommendations could be attributed to random statistical errors. This hypothesis was later used to inform the interpretation of the workshop results, presented under the heading "Review of Recommendations Made at the *Standards* Validation."

Table E-1 shows the median recommended cut scores from the standards validation workshop plus the associated impact data for ELA and mathematics using Spring 2018 administration data. Impact data are the percentages of students who would be classified in each performance level if the cut scores were applied to students' scores. Note that the impact data presented in this document are based on the test data available at the time of the standards validation, so they should not be considered final; however, these impact data provide a reasonable estimate of the percentages of students that would be included in each performance level based on the cut scores shown.

Table E-1a. Cut Scores Associated with Participants' Median Recommendations

Content	Grade	Partially Proficient	Proficient	Advanced
ELA	3	1279	1299.5	1316
ELA	4	1382	1399.5	1417
ELA	5	1481	1499.5	1521
ELA	6	1578	1599.5	1624
ELA	7	1679	1699.5	1726
ELA	8	1775	1794	1828
Math	3	1281	1299.5	1321
Math	4	1376	1397	1417
Math	5	1475	1496	1515
Math	6	1579	1599.5	1614
Math	7	1679	1699.5	1715
Math	8	1777	1799.5	1815

Table E-1b. Impact Data Associated with Participants' Median Recommendations

Content	Grade	Not Proficient	Partially Proficient	Proficient	Advanced
ELA	3	29.70%	25.80%	21.10%	23.30%
ELA	4	32.60%	22.30%	21.50%	23.60%
ELA	5	32.10%	21.20%	25.30%	21.40%
ELA	6	31.30%	27.30%	28.20%	13.20%
ELA	7	29.20%	27.30%	30.70%	12.80%
ELA	8	27.50%	22.50%	38.20%	11.80%
Math	3	27.80%	26.40%	27.20%	18.60%
Math	4	24.70%	28.70%	26.90%	19.70%
Math	5	33.50%	26.60%	23.30%	16.60%
Math	6	34.40%	30.90%	18.60%	16.00%
Math	7	36.20%	28.00%	18.30%	17.50%
Math	8	36.50%	30.80%	14.90%	17.80%

Appendix E-5 Review of the Recommendations Made at the Standards Validation

As hypothesized, educators at the content-based standards validation workshop recommended cut scores that were similar to the existing cut scores. MDE and DRC evaluated the recommendations in context. Table E-2 shows the difference between the median cut score recommendations and the existing cut scores, expressed in multiples of the conditional standard error of measurement (CSEM). The CSEM quantifies the amount of statistical error associated with the test. If a student were tested many times, one would expect her scores to fall within a range of ±1.0 CSEM about 2/3 of the time.

Figures E-1 and E-2 show a graphical representation of the existing cut scores beside the recommended cut scores and their associated CSEM.

Table E-2a. Median Cut Score Recommendations from the Standards Validation, Existing ELA and Math Cut Scores, and Differences in Terms of Conditional Standard Error of Measurement (CSEM)

Content	Grade	Partially Proficient	Proficient	Advanced
ELA	3	1279	1299.5	1316
ELA	4	1382	1399.5	1417
ELA	5	1481	1499.5	1521
ELA	6	1578	1599.5	1624
ELA	7	1679	1699.5	1726
ELA	8	1775	1794	1828
Math	3	1281	1299.5	1321
Math	4	1376	1397	1417
Math	5	1475	1496	1515
Math	6	1579	1599.5	1614
Math	7	1679	1699.5	1715
Math	8	1777	1799.5	1815

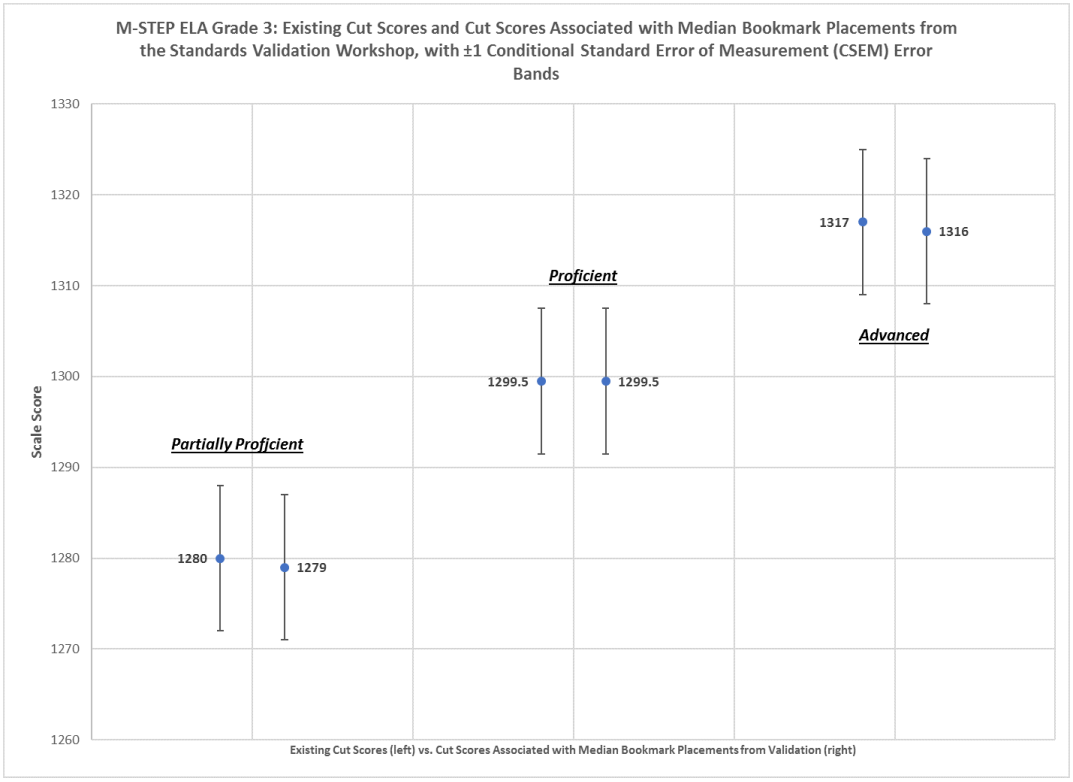
Table E-2b. Median ELA and Math Cut Scores

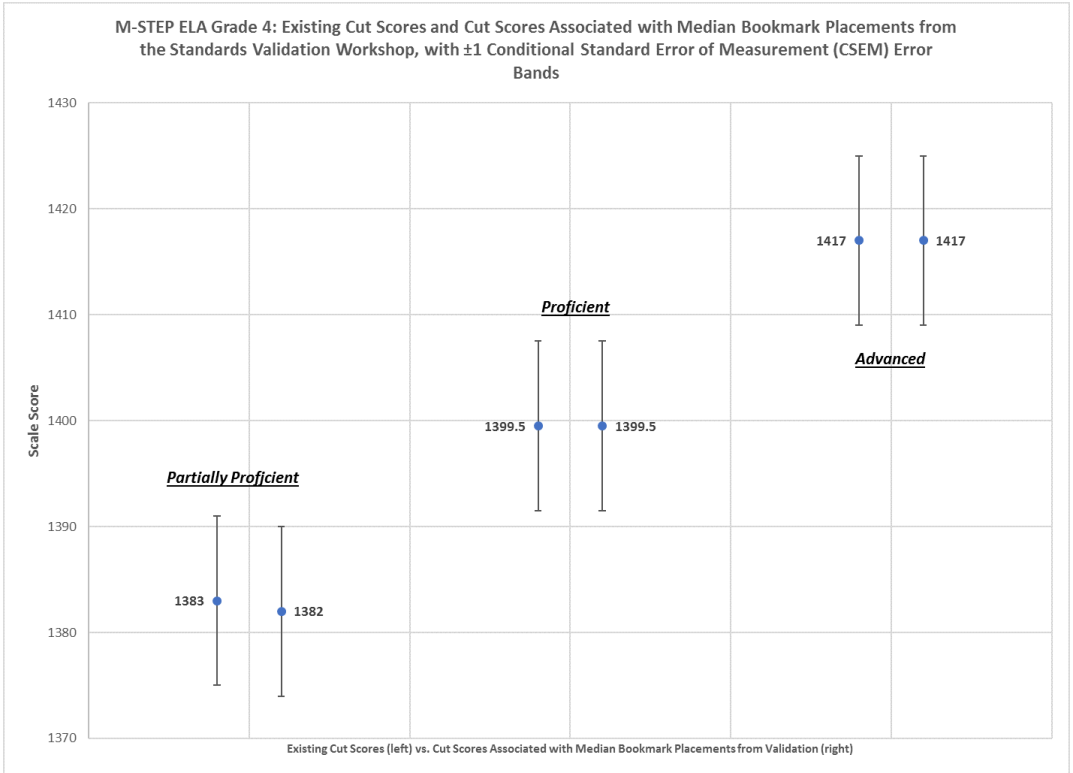
Content	Grade	Partially Proficient	Proficient	Advanced
ELA	3	1280	1299.5	1317
ELA	4	1383	1399.5	1417
ELA	5	1481	1499.5	1524
ELA	6	1578	1599.5	1624
ELA	7	1679	1699.5	1726
ELA	8	1777	1799.5	1828
Math	3	1281	1299.5	1321
Math	4	1376	1399.5	1420
Math	5	1478	1499.5	1515
Math	6	1579	1599.5	1614
Math	7	1679	1699.5	1716
Math	8	1780	1799.5	1815

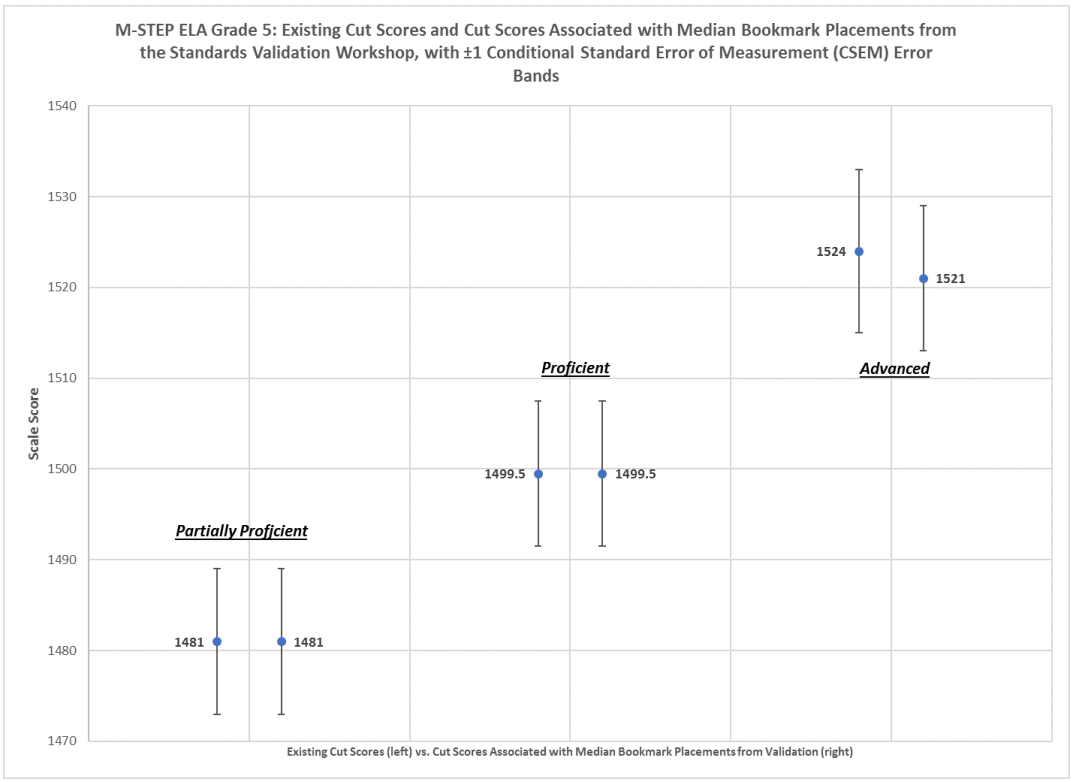
Table E-2c. Differences between Existing and Recommended Cut Scores in Terms of Conditional Standard Error of Measurement (CSEM)

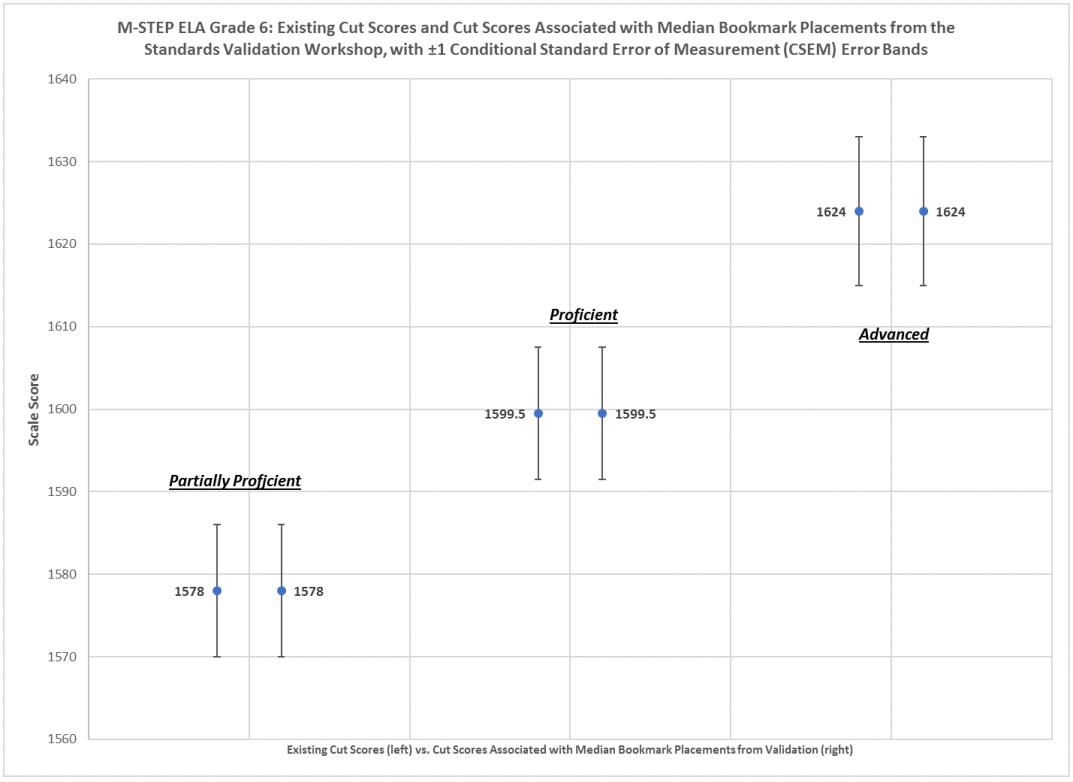
Content	Grade	Partially Proficient	Proficient	Advanced
ELA	3	-0.13	0	-0.13
ELA	4	-0.13	0	0
ELA	5	0	0	-0.38
ELA	6	0	0	0
ELA	7	0	0	0
ELA	8	-0.22	-0.69	0
Math	3	0	0	0
Math	4	0	-0.42	-0.43
Math	5	-0.33	-0.44	0
Math	6	0	0	0
Math	7	0	0	-0.17
Math	8	-0.33	0	0

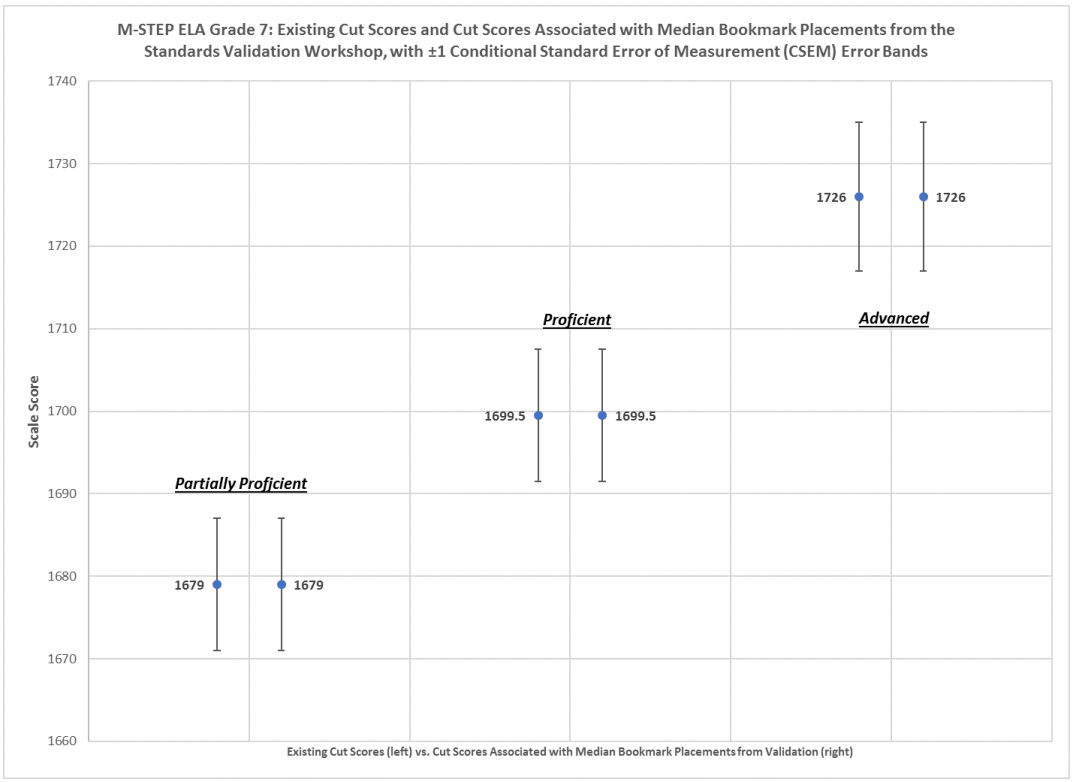
Figure E-1. ELA Comparison of Median Cut Score Recommendations and Existing Cut Scores, with Differences Expressed in Terms of Conditional Standard Error of Measurement (CSEM)






M-STEP ELA Grade 5: Existing Cut Scores and Cut Scores Associated with Median Bookmark Placements from the Standards Validation Workshop, with ±1 Conditional Standard Error of Measurement (CSEM) Error Bands
Scale Score
1540
1530
1520
1510
1500
1490
1480
1470
Partially Proficient
Proficient
Advanced
1481
1481
1499.5
1499.5
1524
1521
Existing Cut Scores (left) vs. Cut Scores Associated with Median Bookmark Placements from Validation (right)


M-STEP ELA Grade 6: Existing Cut Scores and Cut Scores Associated with Median Bookmark Placements from the Standards Validation Workshop, with ±1 Conditional Standard Error of Measurement (CSEM) Error Bands
Scale Score
1640
1630
1620
1610
1600
1590
1580
1570
1560
Partially Proficient
Proficient
Advanced
1578
1578
1599.5
1599.5
1624
1624
Existing Cut Scores (left) vs. Cut Scores Associated with Median Bookmark Placements from Validation (right)


M-STEP ELA Grade 7: Existing Cut Scores and Cut Scores Associated with Median Bookmark Placements from the Standards Validation Workshop, with ±1 Conditional Standard Error of Measurement (CSEM) Error Bands
Scale Score
1740
1730
1720
1710
1700
1690
1680
1670
1660
Partially Proficient
Proficient
Advanced
1679
1679
1699.5
1699.5
1726
1726
Existing Cut Scores (left) vs. Cut Scores Associated with Median Bookmark Placements from Validation (right)

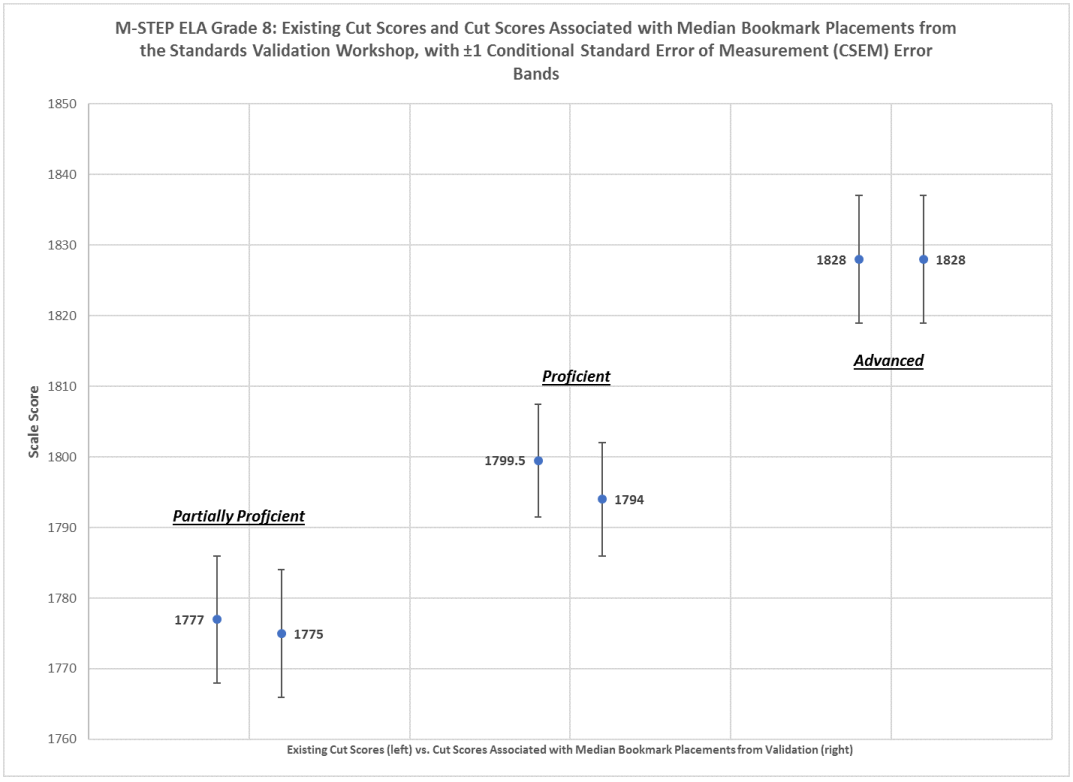
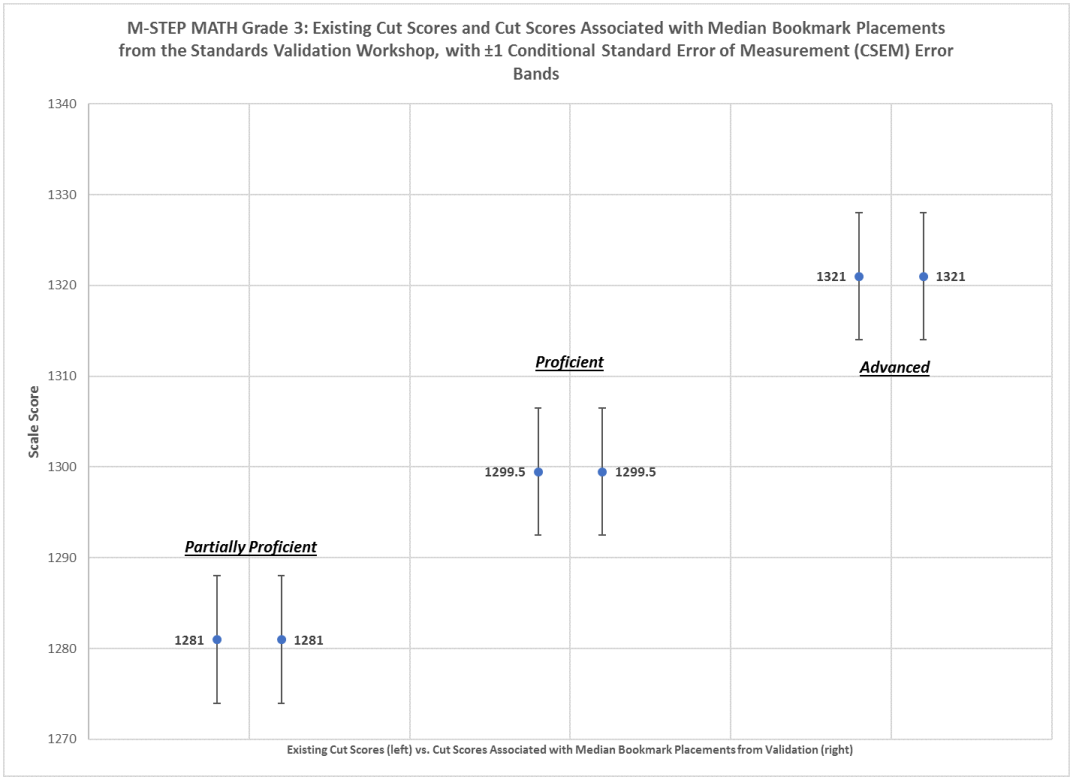
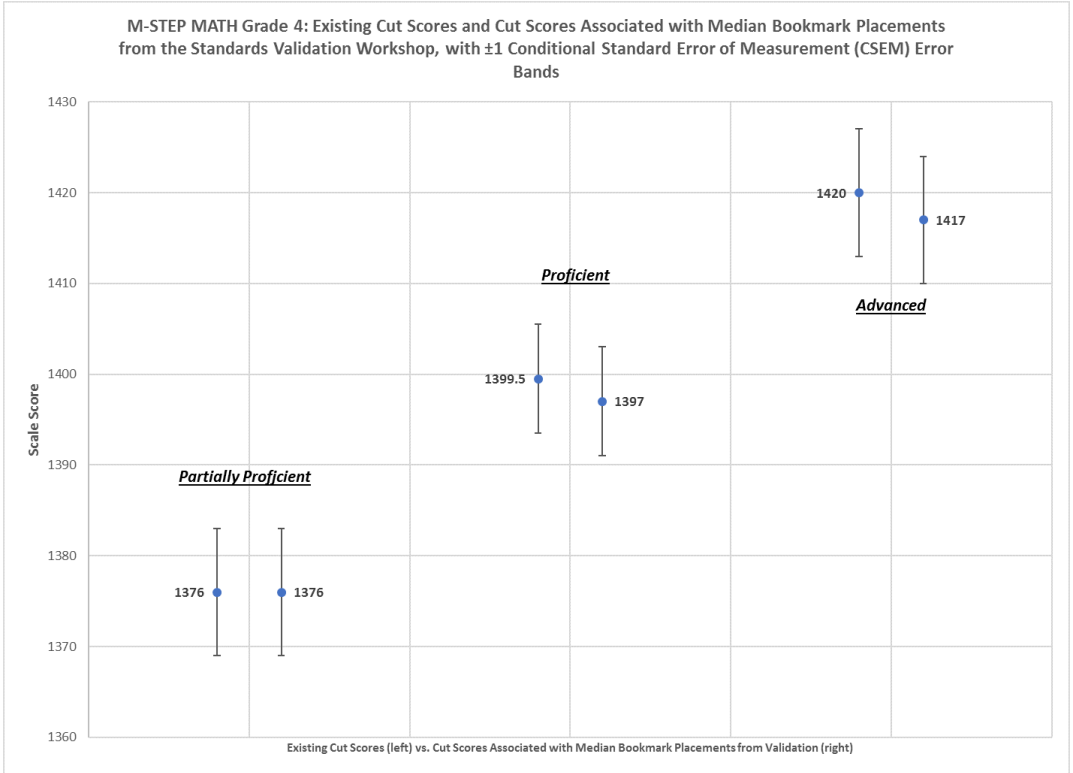
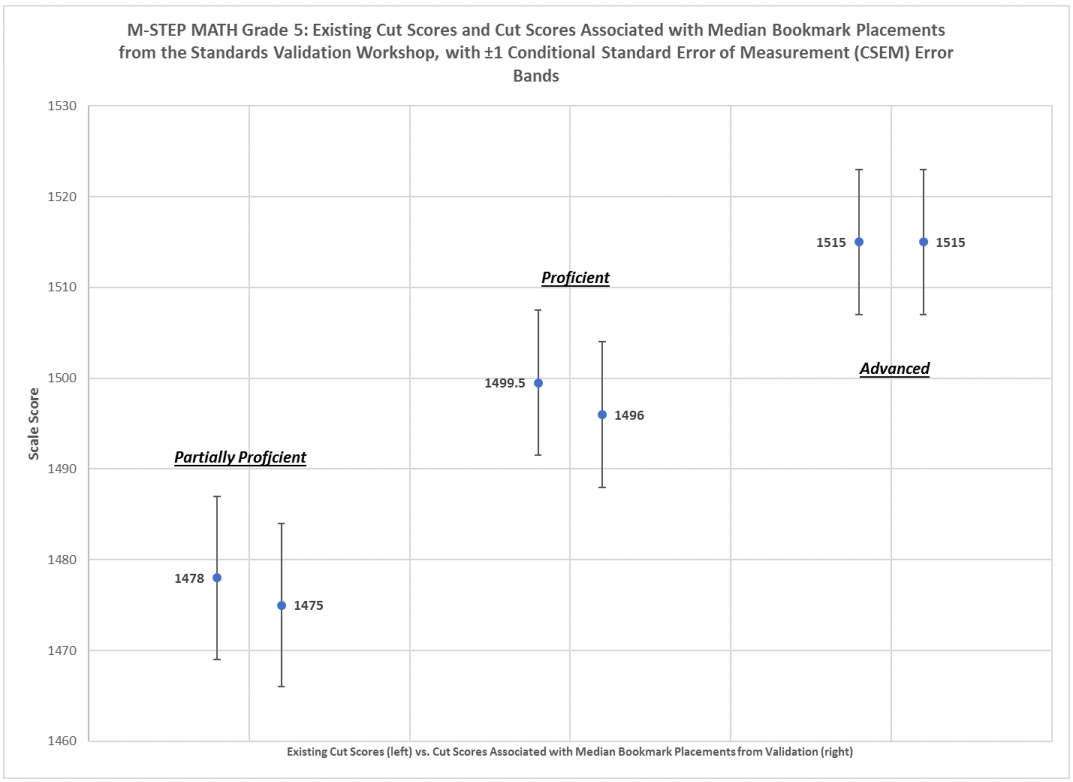
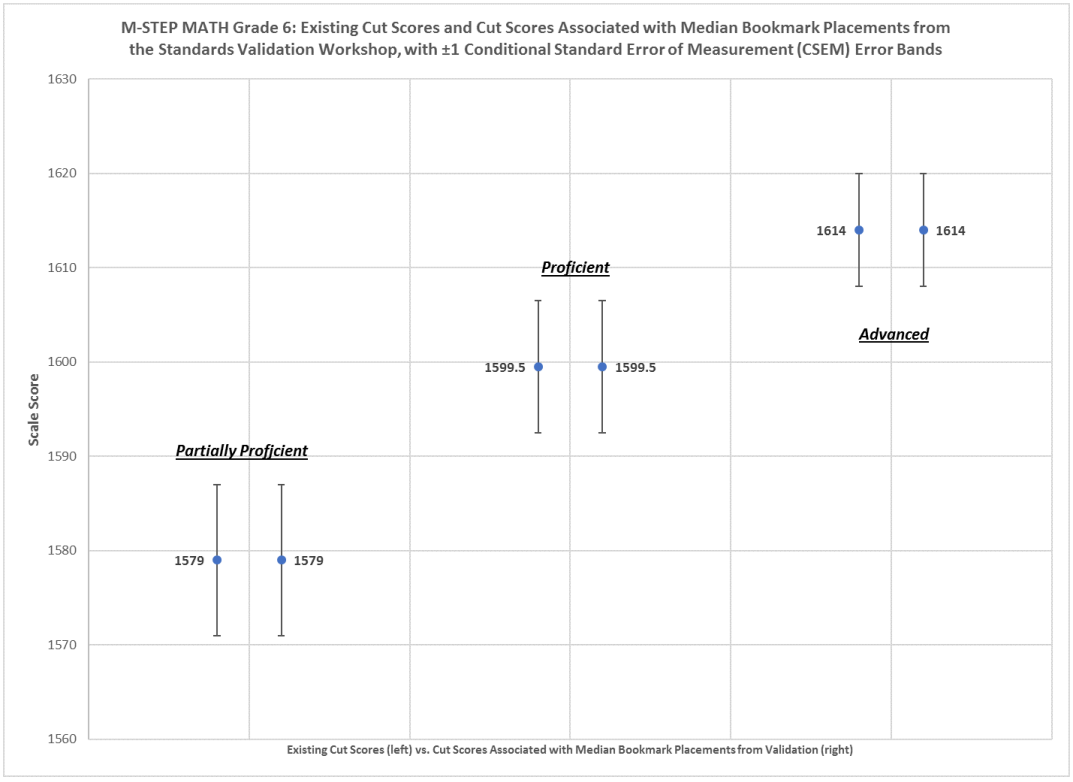

M-STEP ELA Grade 8: Existing Cut Scores and Cut Scores Associated with Median Bookmark Placements from the Standards Validation Workshop, with ±1 Conditional Standard Error of Measurement (CSEM) Error Bands
Scale Score
1850
1840
1830
1820
1810
1800
1790
1780
1770
1760
Partially Proficient
Proficient
Advanced
1777
1775
1799.5
1794
1828
1828
Existing Cut Scores (left) vs. Cut Scores Associated with Median Bookmark Placements from Validation (right)

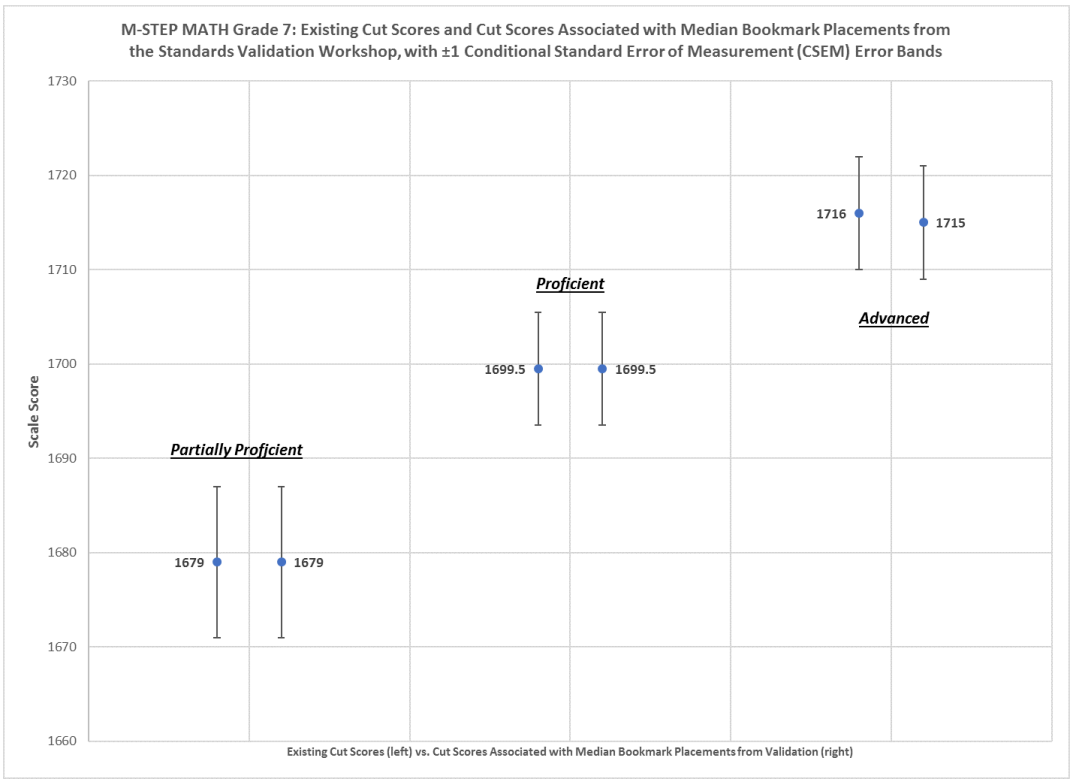
Figure E-2. Mathematics Comparison of Median Cut Score Recommendations and Existing Cut Scores, with Differences Expressed in Terms of Conditional Standard Error of Measurement (CSEM)

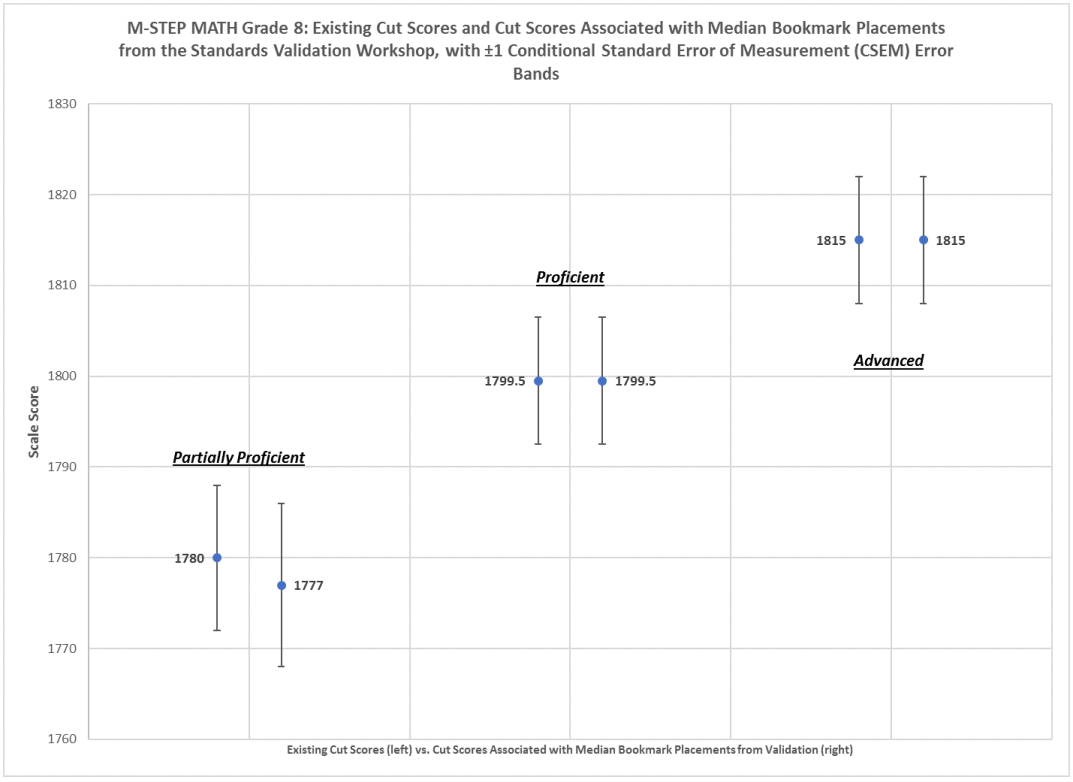





M-STEP MATH Grade 5: Existing Cut Scores and Cut Scores Associated with Median Bookmark Placements from the Standards Validation Workshop, with ±1 Conditional Standard Error of Measurement (CSEM) Error Bands
Scale Score
1530
1520
1510
1500
1490
1480
1470
1460
Partially Proficient
Proficient
Advanced
1478
1475
1499.5
1496
1515
1515
Existing Cut Scores (left) vs. Cut Scores Associated with Median Bookmark Placements from Validation (right)


M-STEP MATH Grade 6: Existing Cut Scores and Cut Scores Associated with Median Bookmark Placements from the Standards Validation Workshop, with ±1 Conditional Standard Error of Measurement (CSEM) Error Bands
Scale Score
1630
1620
1610
1600
1590
1580
1570
1560
Partially Proficient
Proficient
Advanced
1579
1579
1599.5
1599.5
1614
1614
Existing Cut Scores (left) vs. Cut Scores Associated with Median Bookmark Placements from Validation (right)


M-STEP MATH Grade 7: Existing Cut Scores and Cut Scores Associated with Median Bookmark Placements from the Standards Validation Workshop, with ±1 Conditional Standard Error of Measurement (CSEM) Error Bands
Scale Score
1730
1720
1710
1700
1690
1680
1670
1660
Partially Proficient
1679
1679
Proficient
1699.5
1699.5
Advanced
1716
1715
Existing Cut Scores (left) vs. Cut Scores Associated with Median Bookmark Placements from Validation (right)


M-STEP MATH Grade 8: Existing Cut Scores and Cut Scores Associated with Median Bookmark Placements from the Standards Validation Workshop, with ±1 Conditional Standard Error of Measurement (CSEM) Error Bands
Scale Score
1830
1820
1810
1800
1790
1780
1770
1760
Partially Proficient
1780
1777
Proficient
1799.5
1799.5
Advanced
1815
1815
Existing Cut Scores (left) vs. Cut Scores Associated with Median Bookmark Placements from Validation (right)

The MDE considered the recommendations made by the standards validation committee and the existing cut scores. Working with DRC, MDE made three primary findings:

1. The content-based expectations for students in each performance level have not changed significantly since the cut scores were established. Although the tests are now shorter and passage-based writing items have been introduced on the ELA tests, the underlying expectations for students in each performance level have not changed.

2. The impact data observed in spring 2018 is similar to those from the 2017 administration of the tests when the existing cut scores were applied. This similarity supports the contention that the expectations for students in each performance level have not changed, and that the existing cut scores are valid for continued use.

3. The median cut score recommendations were all very close to the existing cut scores, to the point of being statistically indistinguishable. As shown in Table E-2, the average difference from the existing cut scores was -0.11 CSEM, and all were within a range of ±0.7 CSEM. Within this narrow range, it is difficult to argue that scale scores are significantly different.

The available validity evidence suggests that there were no significant differences between the updated ELA and mathematics assessments and the content assessed by the prior assessments; and that the differences between the judgments made at the 2018 standards validation workshop and the existing cut scores were not statistically different. That is, the recommendations made by Michigan educators during the standards validation were consistent with the existing cut scores, and the validity evidence collected during this process supports the continued use of the cut scores.

Table E-3 shows the existing cut scores and associated impact data for ELA and mathematics using spring 2018 administration data. Figures E-3 and E-4 show a graphical representation of the existing cut scores and their associated impact data from spring 2018.

Table E-3a. Existing ELA and Mathematics Cut Scores

Content	Grade	Partially Proficient	Proficient	Advanced
ELA	3	1280	1299.5	1317
ELA	4	1383	1399.5	1417
ELA	5	1481	1499.5	1524
ELA	6	1578	1599.5	1624
ELA	7	1679	1699.5	1726
ELA	8	1777	1799.5	1828
Math	3	1281	1299.5	1321
Math	4	1376	1399.5	1420
Math	5	1478	1499.5	1515
Math	6	1579	1599.5	1614
Math	7	1679	1699.5	1716
Math	8	1780	1799.5	1815

Table E-3b. Associated Impact Data for M-STEP Spring 2018

Content	Grade	Not Proficient	Partially Proficient	Proficient	Advanced
ELA	3	30.90%	24.60%	22.40%	22.10%
ELA	4	33.80%	21.10%	21.50%	23.60%
ELA	5	32.10%	21.20%	28.70%	17.90%
ELA	6	31.30%	27.30%	28.20%	13.20%
ELA	7	29.20%	27.30%	30.70%	12.80%
ELA	8	29.90%	27.30%	31.10%	11.80%
Math	3	27.80%	26.40%	27.20%	18.60%
Math	4	24.70%	33.20%	25.70%	16.40%
Math	5	37.00%	28.50%	17.80%	16.60%
Math	6	34.40%	30.90%	18.60%	16.00%
Math	7	36.20%	28.00%	19.50%	16.40%
Math	8	40.90%	26.30%	14.90%	17.80%

Figure E-3. Existing, Validated Cut Scores and Associated Impact Data for Spring 2018 ELA

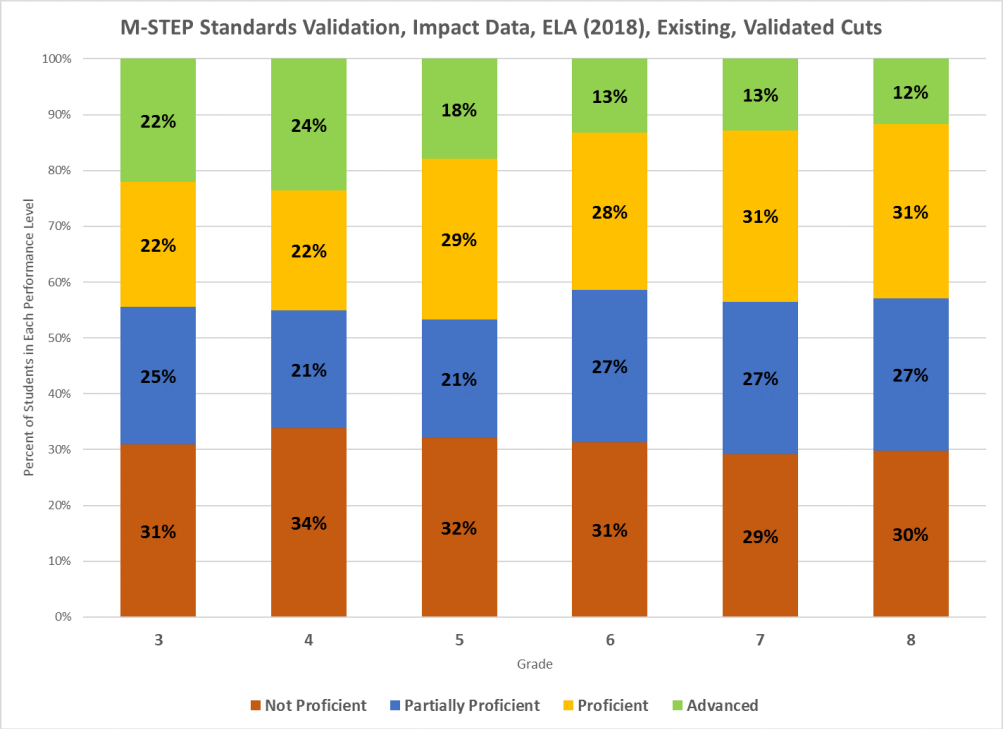
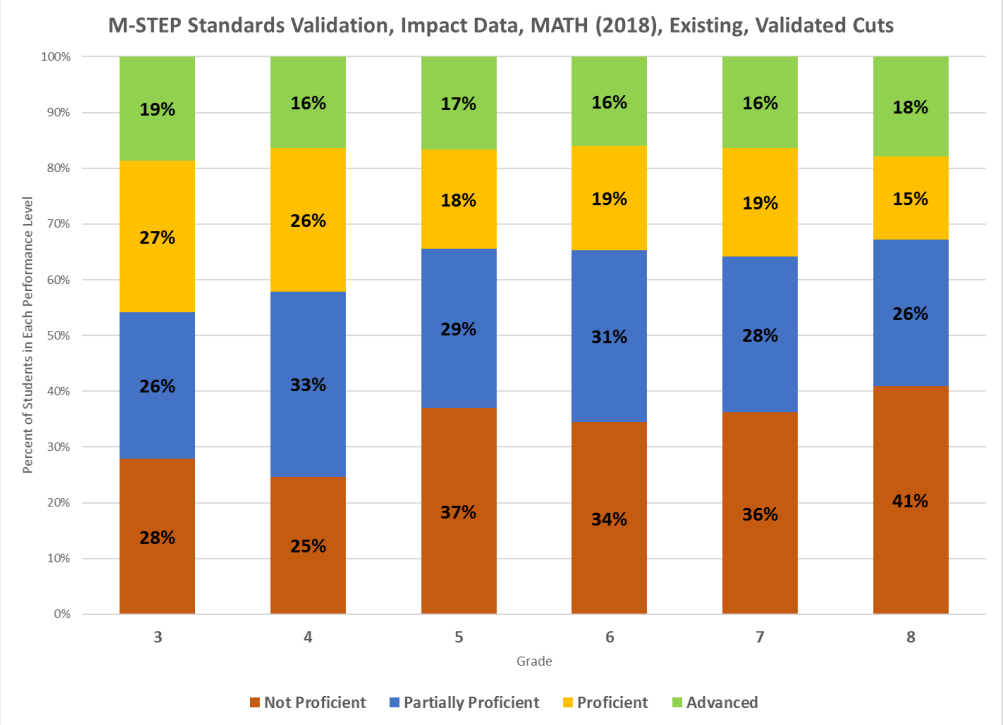


Figure E-4. Existing, Validated Cut Scores and Associated Impact Data for Spring 2018 Mathematics



Appendix E.6 References

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). *Standard setting: A Bookmark approach. Symposium* presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment: Phoenix, AZ.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 225-253). New York: Routledge.

Appendix F: Michigan Assessment System Participant Groups

This appendix provides more details on the stake holders and participants involved in the Michigan Assessment System.

Appendix F.1 Michigan Educators

Michigan educators (including classroom teachers from K–12 and higher education, curriculum specialists, and administrators) play a vital role in all phases of the test development process. Committees of Michigan educators review the test specifications and provide advice on the model or structure for assessing each content area. They also work to ensure that test content and question types align closely with best practices in classroom instruction.

Appendix F.2 Technical Advisory Committee

Michigan's Technical Advisory Committee (TAC) serves as an advisory body to MDE. The TAC provides recommendations on technical aspects of large-scale assessments, including item development, test construction, administration procedures, scoring and equating methodologies, and standard-setting workshops. The TAC also provides guidance on other technical matters, such as practices not already described in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), and continues to provide advice and consultation on the implementation of new assessments and adherence to the federal requirements set forth by the Every Student Succeeds Act. Table F-1 can be referenced for TAC member information.

Table F-1. Technical Advisory Committee

Name	Position	Organization
Dr. Mark Reckase, Chair	Distinguished Professor of Measurement and Quantitative Methods (retired)	Michigan State University
Dr. Damian Betebenner	Senior Associate	National Center for the Improvement of Educational Assessment
Dr. Gregory J. Cizek	Distinguished Professor of Educational Measurement and Evaluation	University of North Carolina, Chapel Hill
Dr. George E. Engelhard, Jr.	Professor Emeritus of Educational Measurement and Policy	University of Georgia
Dr. Christine Carrino Gorowara	Interim Director	Delaware Center for Teacher Education, University of Delaware
Dr. Joseph Martineau	Senior Associate	National Center for the Improvement of Educational Assessment
Dr. Dave Treder	Coordinator of Research, Evaluation, and Assessment	Genesee Intermediate School District, Flint, Michigan

Appendix F.3 Michigan's Division of Educator, Student, and School Supports (DESSS) Advisory Committee

The DESSS Advisory Committee meets quarterly to provide input, ideas, expert advice, and/or recommendations to MDE and DESSS on matters related to assessment and accountability, professional preparation, educator evaluations, assessment policy, and related communications to the field. The committee also meets to keep its respective organizations abreast of changes to the above areas that will affect Michigan's schools and students. The committee comprises representatives from educational agencies, organizations, and representatives from both two-year and four-year colleges and universities across the state. Table F-2 shows the members of the DESSS Advisory Committee.

Table F-2. Division of Educator, Student, and School Supports Advisory Committee

Last Name	First Name	Organization
Anand	Johanna	Michigan Department of Education/Low Incidence Outreach
Arnswald	Jennifer	Michigan Science Teachers Association
Berry	Kathy	Michigan Council of Teachers of Mathematics
Clingman	Cindy	Michigan Reading Association
Cox	Mary	Michigan Council of Teachers of English
Czerwinski	Harvey	Michigan Education Research Association
Dewsbury-White	Kathryn	Michigan Assessment Consortium
DeYoung	Ann	Michigan Elementary and Middle School Principals Association
Flukes	Jonathan	Michigan Education Research Association
Gordon	Casey	MI Council of Teachers of English to Speakers of Other Languages
Greer	Doug	Oakland Area Intermediate School District
Kher	Neelam	Michigan State University
Koekkoek	Matthew	Michigan Association of Administrators of Special Education
Langdon	Thomas	Michigan Association of School Administrators
Mastie	Marge	Washtenaw Intermediate School District - Retired
McIntyre	Rebecca	Michigan Association of Administrators of Special Education
Miller	Kathy	Michigan School Facilitators Network
Trout	Kelly	Ingham Intermediate School District
Vespremi	Stacy	Michigan Association of State and Federal Programs Specialists
Vorenkamp	Ellen	Wayne Regional Educational Services Agency
Zdeb	Wendy	Michigan Association of Secondary School Principals
Substitutes		
McGoran	Holly	Michigan Science Teachers Association
Musial	Joe	Wayne Regional Educational Services Agency
Ripmaster	Colin	Michigan Association of Secondary School Principals
Taraskiewicz	Cindy	Wayne Regional Educational Services Agency