MAP Growth Universal Screening Benchmarks: Establishing MAP Growth as an Effective Universal Screener

March 12, 2021

Wei He, NWEA Psychometric Solutions
Patrick Meyer, NWEA Psychometric Solutions

Acknowledgements: The authors would like to thank the following colleagues for their contributions: Ann Hu and Sarah Tran for linking the MAP Growth with state test scores, Shudong Wang for QA'ing the study and reviewing an earlier version of the report, and Kelly Rivard for copy editing the report.

Suggested citation: He, W., & Meyer, J. (2021). *MAP Growth universal screening benchmarks: Establishing MAP Growth as an effective universal screener*. NWEA.

Table of Contents

List of Tables

Revision History

Date	Version	Description
10/5/2020	0.1	Initial draft created by Wei He
3/12/2021	1.0	Finalized by Patrick Meyer; published

Executive Summary

Universal screening is paramount in identifying students at risk for academic difficulty in a response to intervention (RTI) model, the core of which is to provide students multi-tiered support based on the level of academic risk that students encounter. Typically, a multitiered support system consists of three levels (Tier 1, Tier 2, and Tier 3) representing no intervention needed to the most intense level of intervention. It is estimated that 5–10% of the student population needs the most intensive intervention.

One primary component in RTI is assessment. A universal screening assessment in a particular content domain is typically administered multiple times a year. If a student scores below an established benchmark for a given time point, they are considered to be at risk for learning difficulties in that content domain and in need of intervention. For an assessment to be an effective universal screener, aside from the technical adequacy, it is important and imperative to establish benchmarks through a scientifically designed and evidenced-based process. The benchmarks also need to be explicit as to what level of academic risk they are established to identify (e.g., at some risk or at substantial risk).

NWEA research over the past decade demonstrates the efficacy of using MAP Growth as a universal screening tool. This research and its supporting evidence follow guidelines from the National Center on Intensive Intervention (NCII) in their rating rubrics that delineate technical standards (NCII, 2020a) and their call for submission that provides criteria for submitting evidenced-based universal screening tools (NCII, 2020b). These NCII guidelines are not static across years, and MAP Growth assessments change in ways that require new research and supporting evidence. The NWEA research on universal screening regularly gets updated based on these changes. Most recently, the 2020 MAP Growth norms were released in July 2020 (Thum & Kuhfeld, 2020), which required new research on the efficacy of MAP Growth as a universal screener. Thus, this study was conducted to update MAP Growth universal screener cut scores and provide evidence of their effectiveness.

Specifically, this report documents the process NWEA followed to determine and validate the cut scores for fall, winter, and spring that can be used to identify students in Grades K–8 who have severe learning difficulties and need intensive intervention in reading and mathematics. Universal screening cut scores were first identified and validated for the English MAP Growth Reading and Mathematics assessments, followed by establishing the universal screening cut scores for Spanish MAP Growth Reading.

To establish the universal screening cut scores for the English MAP Growth assessments, the NCII rating rubrics (NCII, 2020a) were followed using a primary sample consisting of students in Arkansas, Colorado, Florida, Missouri and New York and a secondary sample used for cross-validation that consisted of students in Indiana. The primary sample took state-level summative tests and MAP Growth in Spring 2018, whereas the secondary sample took the state summative test and MAP Growth in Spring 2019. While the original Indiana state assessment scale scores (INSS) were used as the criterion measure in the classification accuracy analyses for the secondary sample, state assessment scores from the primary sample were put on the same scale (i.e., the Rasch Unit (RIT) scale) by subject and grade using the equipercentile method to create a common criterion measure and allow state-level test scores to be comparable across states. As a result, each student in the primary sample obtained a MAP Growth linked state score (LSS) in reading and/or mathematics.

The classification accuracy analyses for each sample involved different combinations of candidate cut scores for MAP Growth and different candidate cut scores for the criterion measure for each subject, grade, and term to identify a combination that produced the optimal benchmarks for identifying students in need of intensive intervention in a grade, term, and subject. The classification accuracy analyses results suggest the benchmarks be set at the 30th percentile in MAP Growth Reading and Mathematics for Grades K–8, as shown in Table E.1. Students who score below those benchmarks are likely at risk for severe learning difficulty and in need of intensive intervention. These benchmarks result in sensitivity, specificity, and lower bound of the area under curve (AUC) being at least 0.8, the highest level of the evaluation criteria described in the NCII rating rubrics (NCII, 2020a). The cross-validation study results were consistent with those from the primary sample, providing evidence that the recommended universal screening cut score are valid.

Spanish cut scores were then established by linking the Spanish MAP Growth Reading pilot test scores and the English MAP Growth Reading scores and finding the score on the Spanish assessment that corresponded to the recommended cut scores for the English assessment. The recommended universal screening cut scores for Spanish reading correspond to the 40th percentile of the Spanish MAP Growth Reading norms, as shown in Table E.1. Students with Spanish MAP Growth Reading scores below these cut scores are likely at risk for having difficulty in a subject.

Table E.1. MAP Growth Universal Screening Cut Scores

		MAP Growth Mathematics		MAP Growth Reading		Spanish MAP Growth Reading	
Grade	Term	Cut Score	Percentile	Cut Score	Percentile	Cut Score	Percentile
K	Fall	133	30	130	30	130	40
	Winter	144	30	140	30	140	40
	Spring	151	30	147	30	148	40
1	Fall	154	30	149	30	145	40
	Winter	164	30	159	30	155	40
	Spring	169	30	164	30	163	40
2	Fall	168	30	164	30	165	40
	Winter	177	30	173	30	173	40
	Spring	182	30	177	30	180	40
3	Fall	181	30	178	30	179	40
	Winter	189	30	185	30	185	40
	Spring	194	30	189	30	186	40
4	Fall	192	30	188	30	187	40
	Winter	198	30	194	30	192	40
	Spring	202	30	196	30	195	40
5	Fall	201	30	196	30	194	40
	Winter	206	30	201	30	198	40
	Spring	210	30	203	30	201	40
6	Fall	206	30	202	30	200	40
	Winter	211	30	205	30	202	40
	Spring	214	30	207	30	207	40

Grade	Term	MAP Growth Mathematics		MAP Growth Reading		Spanish MAP Growth Reading	
		Cut Score	Percentile	Cut Score	Percentile	Cut Score	Percentile
7	Fall	211	30	206	30	204	40
	Winter	215	30	209	30	206	40
	Spring	217	30	210	30	211	40
8	Fall	215	30	209	30	207	40
	Winter	218	30	212	30	209	40
	Spring	220	30	213	30	213	40

Please note: The cut scores recommended in this report serve the purpose of identifying students in need of intensive intervention and should not be confused with cut scores used for Read by Grade 3 purposes. The purpose of Read by Grade 3 policies is to identify students who are not on track for proficiency in reading by Grade 3. Read by Grade 3 cut scores are tied to state proficiency standards and correspond to much higher percentiles than universal screening cut scores.

1. Introduction

MAP® Growth™ can be used as a universal screener to identify students who have severe learning difficulties and may need intensive intervention, as evidenced by research NWEA has been conducting for the past decade on the efficacy of using MAP Growth results for universal screening. To conduct this research and establish the MAP Growth universal screening benchmarks (i.e., cut scores), NWEA follows guidelines from the National Center on Intensive Intervention (NCII, 2020a, 2020b) that evolve over time. The universal screening cut scores are updated as needed based on changes to these NCII guidelines and/or the MAP Growth assessments. Most recently, the 2020 MAP Growth norms were released in July 2020 (Thum & Kuhfeld, 2020), which required new research on the efficacy of MAP Growth as a universal screener, hence the purpose of this study to update the MAP Growth universal screening cut scores and provide evidence of their effectiveness.

Specifically, this document presents the universal screening cut scores on MAP Growth Reading and Mathematics from fall, winter, and spring that can be used to identify students in Grades K–8 who have severe learning difficulties and may need intensive intervention. It also presents the universal screening cut scores for Spanish MAP Growth Reading. More importantly, it documents the process NWEA followed to arrive at the cut scores as a source of evidence that they were established through a scientifically designed and evidence-based process, which is imperative for an assessment to be an effective universal screener.

The English MAP Growth cut scores were established first. Once the candidate cut scores were identified based on the MAP Growth norms, classification accuracy analyses were conducted to arrive at the recommended cut scores by evaluating the effectiveness of the cut scores on MAP Growth, depending on the choice of different percentages of students identified as actually in need of intensive intervention based on the state summative test scores (i.e., the criterion measures). Two samples were used in the classification accuracy analyses. The primary sample consisted of students with state summative test scores from five different states (i.e., Arkansas, Colorado, Florida, Missouri, and New York) who took both the state summative and MAP Growth assessments in Spring 2018. The secondary sample consisted of students with scores from the Indiana state summative test administered in Spring 2019. Once the recommended universal screening cut scores were identified for the English assessment, they were used to help establish the universal screening cut scores for Spanish MAP Growth Reading, in conjunction with linking study results between English MAP Growth Reading and Spanish MAP Growth Reading for Grade 3 students.

1.1. MAP Growth Overview

MAP Growth is an adaptive interim assessment aligned to state-specific content standards and designed to measure achievement and growth in Grades K–12 mathematics, reading, language usage, and science. Spanish versions of the MAP Growth tests are also provided for mathematics and reading. Results from research studies actively conducted by NWEA demonstrate the reliability, validity, and fairness of MAP Growth (NWEA, 2020a). MAP Growth scores are reported on the RIT vertical scale that has a mean of 200 and a standard deviation of 10. Scores range between 100 and 350. The RIT scale allows for the measurement of within- and between-year growth in student learning. Scores for each subject area are scaled separately from the other subjects (i.e., the mathematics scale is different from the reading scale).

MAP Growth yields an overall subject area score and instructional area subscores that describe academic strength and weakness in particular areas. MAP Growth is typically administered in the fall, winter, and spring, with an optional summer administration. NWEA periodically conducts a national norming study to produce achievement and growth norms at the individual student level and at the school level. The most recent norming study for MAP Growth was released in July 2020 (Thum & Kuhfeld, 2020).

1.2. Response to Intervention (RTI) Model

The use of a multitiered framework for student support is the core of any response-to-intervention (RTI) model, which aims for early identification of struggling students to give them the support they need to be successful in school. A multitiered support system typically consists of three levels referred to as Tier 1, Tier 2, and Tier 3. Tier 3 is considered the most intense level of intervention (NCII, n.d.). A triangle is often used to depict these three levels of support in an RTI model, with the top portion representing 5–10% of the student population needing the most intensive intervention (Tier 3), the middle representing 5–15% of the student population needing the targeted but less intensive intervention (Tier 2), and the base representing 80–90% of the students receiving no additional (Tier 1).

Assessment is a primary component of RTI and can be classified into two categories: universal screening and progress monitoring. Universal screening identifies at-risk students in need of intervention, whereas progress monitoring tracks the learning progress of students who are already identified as at-risk for learning difficulty. Ysseldyke et al. (2010) provide an example of one well-accepted RTI model for assessing students:

1. Screen all students in the fall, winter, and spring.
2. Identify low achievers and monitor them monthly.
3. Monitor students needing intensive intervention at least weekly.

This model indicates the need and importance of using appropriate assessment in the RTI process. An RTI model cannot sustain itself in the absence of a technically sound assessment system. For example, if the assessment lacks sufficient reliability and validity to measure student performance or if the cut scores used to screen for at-risk students are not well established with adequate classification accuracy, the decision-making will be untrustworthy and the intervention is likely to be ineffective.

1.3. Universal Screening vs. Read by Grade 3 Cut Scores

There is an important distinction about using MAP Growth scores as a universal screener vs. in states' Read by Grade 3 policies. The two uses are very different and are not interchangeable. While MAP Growth scores are also used to identify students who are not on track for proficiency by Grade 3 according to state Read by Grade 3 policies, cut scores established for the Read by Grade 3 policies are based on state proficiency standards and yielded in a separate linking study between MAP Growth Reading and state English language arts (ELA) tests. Proficiency represents a much higher standard for academic performance than universal screening. Consequently, Read by Grade 3 cut scores correspond to high percentiles on MAP Growth than do universal screening cut scores.

2. Universal Screening Cut Scores for MAP Growth Reading and Mathematics

2.1. Student Sample

Two student samples were used in the study. The primary student sample was used as a criterion measure to conduct the classification accuracy analyses, while the secondary student sample was used for cross-validation. Table 2.1 presents the number of students in each study sample across grades by state and subject area, as well as the U.S. census region and division to which each state belongs. The primary study sample, which was considered the "national sample" based on the NCII rating rubrics (NCII, 2020a), included students in Grades K–8 from five states (Arkansas, Colorado, Florida, Missouri, and New York) covering all four U.S. census regions and five census divisions. The secondary sample consisted of Indiana students in Grades K–8. To be included in the study sample, students participated in both the state summative test and MAP Growth during either the Spring 2018 or Spring 2019 administration for the primary and secondary samples, respectively.

MAP Growth scores from fall, winter, and spring within the same academic year were extracted for Grades 3–8 students in both study samples. Students normally do not begin taking their state summative assessment until Grade 3, so longitudinal MAP Growth data were collected for the Grade 3 cohort in each sample group to obtain information for Grades K–2. For example, to accomplish this for the primary sample group, MAP Growth results were used from 2017–2018 for Grade 3, from 2016–2017 for Grade 2, from 2015–2016 for Grade 1, and from 2014–2015 for Grade K.

Table 2.1. Number of Students in Each Study Sample

Sample Group	State	U.S. Census Region	U.S. Census Division	#Students			
				Mathematics		Reading	
				N	%	N	%
Primary	AR	South	West South Central	21,526	14.2	21,366	14.1
	CO	West	Mountain	26,446	17.4	24,345	16.1
	FL	South	South Atlantic	52,755	34.8	56,717	37.4
	MO	Midwest	West North Central	15,323	10.1	14,424	9.5
	NY	Northeast	Middle Atlanta	35,533	23.4	35,750	23.6
	Overall			151,583	100.0	152,602	100.0
Secondary	IN	Midwest	East North Central	241,079	100.0	244,023	100.0

Table 2.2 presents the number of students by grade, race/ethnicity, and sex. For both samples, White students were the dominant group, followed by Hispanic and Black students, respectively, and there were slightly more male students than female students for most grades. Compared with the primary sample group, the secondary group consisted of more White students.

Table 2.2. Study Sample Demographics

Grade	Sample	#Students	%Students by Race/Ethnicity*								%Students by Sex	
			White	Black	Hispanic	Asian	AI	NH	MR	Unknown	Female	Male
Mathematics												
K	Primary	9,856	41.2	12.4	28.2	5.8	1.2	1.8	2.3	7.1	50.3	49.7
	Secondary	15,348	67.9	12.2	13.2	1.6	0.1	0.0	4.9	–	48.6	51.4
1	Primary	13,799	45.3	12.8	27.1	5.1	1.1	1.4	3.1	4.2	49.6	50.4
	Secondary	24,398	69.0	10.7	13.3	1.6	0.1	0.1	5.1	–	48.1	51.9
2	Primary	13,026	43.8	15.3	29.5	3.7	1.3	1.3	3.1	2.0	49.4	50.6
	Secondary	21,347	72.8	9.1	11.1	1.9	0.1	0.1	5.0	–	47.8	52.2
3	Primary	30,828	48.0	14.2	24.3	4.6	0.7	0.8	3.5	3.8	49.0	51.0
	Secondary	40,103	65.6	12.8	14.1	1.9	0.1	0.1	5.2	–	48.2	51.8
4	Primary	29,854	48.0	13.9	24.0	5.1	0.8	0.8	3.4	4.0	49.8	50.2
	Secondary	40,457	66.0	12.6	14.3	1.8	0.2	0.1	5.0	–	49.2	50.8
5	Primary	29,015	47.9	13.2	25.4	4.9	0.8	0.8	3.1	3.9	49.4	50.6
	Secondary	41,410	65.5	12.7	14.7	1.8	0.1	0.1	5.1	–	49.0	51.0
6	Primary	25,066	47.7	13.0	26.5	4.8	1.0	1.0	2.6	3.5	49.4	50.6
	Secondary	40,638	66.6	12.2	14.6	1.5	0.2	0.1	4.8	–	48.9	51.1
7	Primary	21,200	47.8	11.6	26.4	4.8	1.0	1.0	2.5	4.9	49.9	50.1
	Secondary	40,047	66.7	12.5	14.1	1.8	0.2	0.1	4.8	–	49.0	51.0
8	Primary	15,621	49.5	13.3	25.8	3.9	1.1	0.4	2.6	3.6	49.3	50.7
	Secondary	38,438	67.5	12.2	14.0	1.6	0.2	0.1	4.5	–	48.7	51.3
Reading												
K	Primary	8,321	39.3	12.8	32.0	4.7	1.1	2.1	1.6	6.4	50.9	49.1
	Secondary	15,786	68.4	11.8	13.0	1.6	0.1	0.1	4.9	–	48.7	51.3
1	Primary	13,064	46.6	12.3	27.5	4.1	1.0	1.4	2.6	4.4	49.8	50.2
	Secondary	24,652	69.2	10.6	13.3	1.6	0.1	0.1	5.1	–	48.0	52.0
2	Primary	12,846	43.7	15.4	29.8	3.6	1.3	1.3	3.1	1.8	49.7	50.3
	Secondary	22,393	73.4	8.8	10.9	1.8	0.1	0.1	5.0	–	48.5	51.5
3	Primary	30,698	47.8	14.8	24.0	4.6	0.7	0.8	3.6	3.6	49.1	50.9
	Secondary	40,699	65.9	12.7	14.0	1.9	0.1	0.1	5.2	–	48.2	51.8
4	Primary	29,771	48.1	13.7	24.3	5.0	0.8	0.8	3.4	3.9	49.8	50.2
	Secondary	41,109	66.3	12.5	14.2	1.7	0.2	0.1	5.0	–	49.3	50.7
5	Primary	28,388	47.9	13.3	25.5	4.9	0.7	0.9	3.0	3.9	49.5	50.5
	Secondary	41,928	65.9	12.5	14.6	1.7	0.1	0.1	5.1	–	49.1	50.9
6	Primary	24,148	46.9	13.3	26.6	4.8	1.1	0.9	2.5	3.8	49.6	50.4
	Secondary	41,224	66.8	12.2	14.5	1.5	0.2	0.1	4.8	–	48.9	51.1
7	Primary	21,777	47.1	12.2	27.0	5.3	1.0	0.9	2.4	4.0	49.8	50.2
	Secondary	40,209	66.8	12.4	14.0	1.8	0.2	0.1	4.8	–	49.0	51.0
8	Primary	17,820	46.3	14.0	26.2	5.8	1.3	0.3	2.4	3.8	49.3	50.7
	Secondary	38,868	67.8	12.0	13.8	1.6	0.2	0.1	4.5	–	48.9	51.1

*AI = American Indian. NH = Native Hawaiian. MR = Multi-Race.

2.2. Linking the Primary Sample State Test Scores

The primary sample took different state summative assessments with different underlying scales. In order for the scores to be compared across states, they had to be put on the same scale by subject and grade (i.e., Grades 3–8) using the equipercentile method that matches scores from MAP Growth and the state summative assessments through percentile ranks. This linking strategy brought the test scores from each state assessment to a common metric (i.e., the MAP Growth RIT scale). As a result, each student in the primary sample obtained a MAP Growth linked state score in reading and/or mathematics.[1] Linking the state summative test scores to a common metric allowed state-level test scores to be comparable across states, creating a common measure across states that served as the criterion measure known as the linked state scores (LSS) in the descriptive statistics and classification accuracy analysis for the primary sample.[2]

2.3. Descriptive Statistics of Test Scores from the Student Sample

Table 2.3 presents the descriptive statistics of test scores for the primary sample, including the correlation coefficients (*r*) between the LSS and MAP Growth scores for each term, the total sample size, and the means and standard deviations of the LSS and MAP Growth scores for each term. Table 2.4 presents the same sets of statistics for the secondary sample, with the Indiana state assessment scale score (INSS) as the criterion measure.

The correlation coefficients can be viewed as a type of criterion validity evidence that indicate the degree of relationship of performance on two measures in the same domain area. Criterion validity can be further categorized as concurrent and predictive depending on when the state summative and MAP Growth tests were taken. Concurrent validity occurs when the assessments are taken during the same term (i.e., MAP Growth spring vs. state summative spring), whereas predictive validity occurs when taken during different terms (i.e., MAP Growth fall and winter vs. state summative spring).

Strong validity evidence is indicated if the correlations are above 0.8. For Grades 3–8, almost all concurrent validity indices were above 0.8 for both subjects, and a large proportion of predictive validity indices were either above 0.8 or at least at the high end of 0.7. For Grades K–2, most correlation coefficients were between 0.7 and 0.8 but were generally smaller than those for Grades 3–8. The coefficients for Grades K–2 also decreased accordingly as grades went lower. This was expected as the time lapse between the state assessment and MAP Growth K–2 became longer from Grade 2 to Grade K. Compared with reading, mathematics tended to see higher correlation coefficients between the state and MAP Growth test scores for all grades.

[1] For a detailed description of the linking method, please refer to the Indiana MAP Growth linking study report (NWEA, 2020b).

[2] The approach of linking test scores on different state standardized tests to a common scale is documented and validated in a study by Reardon et al. (2017) who transformed state-level test scores to the common national NAEP scale that yields score distributions corresponding well to the relative performance of students in different districts on the NAEP and MAP Growth assessments.

Table 2.3. Descriptive Statistics of Test Scores—Primary Sample

	r (LSS, MAP Growth)*			#Students				Mean				SD			
					MAP Growth				MAP Growth				MAP Growth		
Grade	Fall 2017	Winter 2018	Spring 2018	LSS	Fall 2017	Winter 2018	Spring 2018	LSS	Fall 2017	Winter 2018	Spring 2018	LSS	Fall 2017	Winter 2018	Spring 2018
Mathematics															
K	0.58	0.63	0.68	–	9,161	7,120	9,856	–	145.2	153.2	161.0	–	10.6	11.3	9.6
1	0.68	0.72	0.74	–	13,467	12,664	13,799	–	161.2	168.6	175.1	–	10.0	10.2	10.6
2	0.73	0.77	0.78	–	12,196	12,577	13,026	–	176.3	184.1	190.8	–	13.2	12.3	12.4
3	0.78	0.82	*0.84*	30,828	28,903	28,419	30,828	202.1	189.8	196.2	202.1	13.7	12.5	12.3	13.2
4	0.79	0.82	*0.85*	29,854	27,925	27,110	29,854	212.7	201.5	206.5	212.9	15.0	12.7	12.9	14.4
5	0.81	0.83	*0.86*	29,015	27,496	26,478	29,015	221.7	211.8	215.9	221.7	16.6	14.3	14.6	16.4
6	0.81	0.83	*0.85*	25,066	22,907	22,299	25,066	223.4	215.3	218.8	223.2	17.0	14.6	15.0	16.4
7	0.83	0.84	*0.86*	21,200	19,513	18,341	21,200	227.7	221.5	224.2	227.9	17.8	15.8	16.1	17.5
8	0.82	0.83	*0.84*	15,621	14,220	13,320	15,621	229.6	225.1	228.0	229.8	19.3	17.3	17.5	18.7
Reading															
K	0.52	0.58	0.62	–	7,751	6,958	8,321	–	140.6	149.2	154.4	–	9.7	11.1	11.1
1	0.64	0.69	0.72	–	12,788	12,293	13,064	–	154.6	162.6	169.2	–	11.7	12.6	12.5
2	0.72	0.76	0.77	–	11,862	12,308	12,846	–	174.3	183.0	189.3	–	17.2	16.7	16.0
3	0.78	0.80	*0.81*	30,698	28,784	28,454	30,698	198.6	188.7	194.4	198.6	15.4	16.2	15.6	15.4
4	0.77	0.79	*0.80*	29,771	27,863	27,249	29,771	206.2	199.1	203.4	206.2	14.7	15.4	14.6	14.7
5	0.78	0.79	*0.81*	28,388	26,879	25,939	28,388	212.3	206.7	210.1	212.4	14.5	15.0	14.3	14.4
6	0.78	0.79	*0.79*	24,148	22,396	21,801	24,148	215.8	211.5	214.1	215.7	14.9	15.3	14.7	15.0
7	0.77	0.78	*0.78*	21,777	20,031	19,206	21,777	219.4	215.8	217.9	219.3	15.4	15.2	14.8	15.2
8	0.76	0.77	*0.77*	17,820	16,417	15,348	17,820	222.1	219.7	221.6	222.0	15.5	15.3	14.8	15.5

*The correlation coefficients for Grades K–2 were computed between the MAP Growth linked state scores for Grade 3 students and MAP Growth scores when Grade 3 students were in Grades K–2. Italicized numbers indicate concurrent validity indices.

Table 2.4. Descriptive Statistics of Test Scores—Secondary Sample

Grade	*r* (INSS, MAP Growth)*			#Students				Mean				SD			
					MAP Growth				MAP Growth				MAP Growth		
	Fall 2018	Winter 2019	Spring 2019	INSS	Fall 2018	Winter 2019	Spring 2019	INSS	Fall 2018	Winter 2019	Spring 2019	INSS	Fall 2018	Winter 2019	Spring 2019
Mathematics															
K	0.61	0.68	0.70	–	12,317	13,009	15,348	–	147.0	154.3	161.9	–	8.7	9.8	10.0
1	0.71	0.75	0.77	–	24,067	23,000	24,398	–	161.7	169.4	176.1	–	10.1	9.9	10.5
2	0.76	0.80	0.81	–	20,235	20,026	21,347	–	177.9	186.9	193.5	–	13.3	12.2	12.5
3	0.83	0.87	*0.89*	40,103	38,266	37,650	40,103	6435.5	189.6	197.0	202.8	76.4	13.3	13.0	13.8
4	0.85	0.88	*0.90*	40,457	38,829	37,908	40,457	6475.3	202.0	207.3	212.9	78.3	13.4	13.4	15.3
5	0.86	0.89	*0.91*	41,410	39,846	39,141	41,410	6497.8	211.2	216.1	221.4	85.1	14.8	15.4	17.5
6	0.87	0.89	*0.90*	40,638	39,078	37,618	40,638	6523.7	217.5	221.6	225.2	93.2	15.1	15.7	16.8
7	0.88	0.89	*0.90*	40,047	38,355	35,618	40,047	6533.7	224.9	227.4	230.8	96.2	16.4	16.9	17.9
8	0.86	0.87	*0.89*	38,438	36,766	34,202	38,438	6548.7	230.5	232.4	235.4	106.3	17.2	17.8	19.0
Reading															
K	0.50	0.59	0.63	–	12,860	13,378	15,786	–	138.9	146.8	154.5	–	8.9	10.1	11.0
1	0.65	0.70	0.72	–	24,556	23,410	24,652	–	154.9	163.1	170.3	–	11.6	12.4	12.5
2	0.73	0.77	0.76	–	21,083	21,421	22,393	–	176.2	186.1	191.6	–	16.9	16.1	15.3
3	0.78	0.80	*0.82*	40,699	38,857	38,275	40,699	5447.0	187.8	194.8	199.1	69.1	16.5	15.8	15.7
4	0.78	0.80	*0.82*	41,109	39,422	38,575	41,109	5477.8	198.8	203.3	205.9	75.1	15.4	14.8	15.4
5	0.79	0.80	*0.81*	41,928	40,396	39,633	41,928	5508.5	205.4	209.0	210.9	79.4	15.4	14.7	15.2
6	0.78	0.79	*0.80*	41,224	39,347	37,965	41,224	5531.4	210.9	213.7	215.3	73.2	15.2	14.6	15.2
7	0.79	0.79	*0.81*	40,209	37,995	35,547	40,209	5557.3	215.6	217.3	218.8	81.4	15.2	15.0	15.2
8	0.79	0.79	*0.81*	38,868	36,806	34,305	38,868	5570.9	219.2	220.6	222.0	77.9	15.2	15.1	15.4

*The correlation coefficients for Grades K–2 were computed between the MAP Growth linked state scores for Grade 3 students and MAP Growth scores when Grade 3 students were in Grades K–2. Italicized numbers indicate concurrent validity indices.

2.4. Candidate MAP Growth Cut Scores

Establishing the benchmark for what constitutes severe learning needs is a key step in an RTI process. While there is no clear census on what should be used to identify students at risk for severe learning needs, a recommended approach is to use national norms for the assessment used for the screening purpose (Crawford, 2014). Because the development of national norms tends to use larger and nationally representative norming samples, they typically provide accurate and reliable information about the relative standing of an individual student against their peers. If a student's score is lower than an established cut scores based on a national norm, this student may require intensive intervention.

Based on research findings from the RTI literature, this study considered the MAP Growth scores in the fall, winter, and spring corresponding to the 10th–40th percentile ranks from the 2020 MAP Growth norms as the candidate cut scores for both reading and mathematics (i.e., 31 sets of candidate MAP Growth cut scores by subject, grade, and term). If a student's MAP Growth score in a term is lower than a given candidate cut score, they were flagged as at-risk in the classification accuracy analysis. Table 2.5 presents the candidate MAP Growth universal screening cut scores at an interval of five percentile ranks by subject, grade, and term.

Table 2.5. Candidate MAP Growth Cut Scores

Grade	Term	Candidate MAP Growth Cut Scores by Percentile Rank						
		10th	15th	20th	25th	30th	35th	40th
Mathematics								
K	Fall	124	127	129	131	133	135	136
	Winter	135	138	140	142	144	146	147
	Spring	142	145	147	149	151	152	154
1	Fall	144	147	150	152	154	155	157
	Winter	154	157	160	162	164	165	167
	Spring	160	163	165	168	169	171	173
2	Fall	158	162	164	166	168	170	172
	Winter	167	171	173	175	177	179	181
	Spring	172	175	178	180	182	184	186
3	Fall	171	175	177	179	181	183	185
	Winter	179	182	185	187	189	191	193
	Spring	183	186	189	192	194	196	198
4	Fall	181	185	187	190	192	194	196
	Winter	187	191	194	196	198	200	202
	Spring	191	194	197	200	202	205	207
5	Fall	190	193	196	199	201	203	205
	Winter	194	198	201	204	206	209	211
	Spring	197	201	205	207	210	212	215
6	Fall	194	198	201	204	206	209	211
	Winter	198	202	205	208	211	213	215
	Spring	200	205	208	211	214	216	218

Grade	Term	Candidate MAP Growth Cut Scores by Percentile Rank						
		10th	15th	20th	25th	30th	35th	40th
7	Fall	198	202	206	208	211	213	216
	Winter	201	205	209	212	215	217	219
	Spring	203	207	211	214	217	220	222
8	Fall	201	205	209	212	215	218	220
	Winter	203	208	212	215	218	221	223
	Spring	205	210	214	217	220	223	225
Reading								
K	Fall	121	124	126	128	130	132	134
	Winter	131	134	136	138	140	142	143
	Spring	138	141	143	145	147	148	150
1	Fall	140	143	145	147	149	151	153
	Winter	149	152	155	157	159	161	163
	Spring	153	157	159	162	164	166	168
2	Fall	153	157	160	162	164	166	168
	Winter	162	166	169	171	173	175	177
	Spring	166	170	173	175	177	180	182
3	Fall	165	169	173	175	178	180	182
	Winter	173	177	180	183	185	188	190
	Spring	176	180	183	186	189	191	193
4	Fall	175	179	183	185	188	190	192
	Winter	182	186	189	192	194	196	198
	Spring	184	188	191	194	196	199	201
5	Fall	183	187	191	193	196	198	200
	Winter	189	193	196	198	201	203	205
	Spring	191	194	198	200	203	205	207
6	Fall	189	193	196	199	202	204	206
	Winter	193	197	200	203	205	208	210
	Spring	195	199	202	205	207	209	211
7	Fall	193	197	200	203	206	208	210
	Winter	196	200	203	206	209	211	213
	Spring	197	201	205	207	210	212	214
8	Fall	196	200	204	207	209	211	214
	Winter	199	203	206	209	212	214	216
	Spring	200	204	207	210	213	215	217

2.5. Candidate Criterion Measure Cut Scores

A criterion measure, also known as an outcome measure, was needed to evaluate the effectiveness of the MAP Growth cut scores to identify students in need of intensive intervention. For this study, scores from the state summative assessments were the criterion measures (i.e., LSS for the primary sample and INSS for the secondary sample). As state-level summative assessments typically start from Grade 3, state assessment scores for Grade 3 students were used as the criterion measure for Grades K–2.

While the cut scores from the primary sample are the result of the equipercentile linking to put the five state assessments onto one scale (i.e., the RIT scale), the cut scores for the secondary sample are the original Indiana state-level scale scores. Following the NCII rating rubrics (NCII, 2020a), students in both the primary and secondary samples who scored at the bottom 10th to 20th percentile ranks based on the state score common metric were designated as "actually at-risk" students (i.e., 11 sets of candidate cut scores were considered for the state-level summative assessment for each subject and grade). If a student's state summative assessment score was lower than a candidate criterion measure cut score, they were identified as actually at-risk in the classification accuracy analysis. Table 2.6 provides the candidate cut scores on the criterion measures, at an interval of 5 percentile ranks, for both the primary and secondary samples.

Table 2.6. Candidate Criterion Measure Cut Scores

	Candidate Criterion Measure Cut Score by Percentile Rank					
	Primary Sample (RIT Scale)			Secondary Sample (IN Scale)		
Grade	10th	15th	20th	10th	15th	20th
Mathematics						
3	185	190	192	6335	6354	6370
4	195	199	202	6374	6394	6410
5	202	206	209	6387	6409	6426
6	203	208	211	6400	6428	6450
7	207	211	214	6407	6433	6453
8	207	211	215	6408	6435	6456
Reading						
3	178	183	186	5357	5373	5387
4	188	192	196	5379	5397	5413
5	195	199	202	5403	5424	5441
6	197	202	205	5436	5456	5470
7	201	205	209	5449	5472	5490
8	203	208	211	5469	5490	5506

2.6. Classification Accuracy Analysis

2.6.1. Overview

The degree to which MAP Growth can accurately identify students who need intensive intervention was evaluated using classification accuracy statistics based on the candidate MAP Growth cut scores that show the proportion of students correctly classified by their RIT scores as at-risk or not-at-risk and the candidate criterion measure cut scores that show whether students actually need intensive intervention. The study considered 31 sets of candidate MAP Growth cut scores and 11 sets of candidate criterion measure cut scores for each subject, grade, and term. This resulted in a total of 341 (31×11) classification accuracy analyses for each subject, grade, and term. Each analysis was conducted as follows:

1. Compare an individual student's (a) MAP Growth score and the candidate MAP Growth cut score and (b) their score on the criterion measure and the criterion measure cut score. Assign "1" in one of the four designations demonstrated in the two-by-two classification table in Table 2.7.
2. Aggregate the designations to obtain the total counts in each cell for students in the sample.
3. Compute the statistics in Table 2.8.

The classification accuracy statistics for the same subject, grade, and term were compared with each other and evaluated against the NCII rating rubrics (NCII, 2020a). A candidate MAP Growth cut score was considered "good" if its sensitivity, specificity, and lower bound of the AUC were all at least 0.8, and it was subsequently recommended as the universal screening cut score for that subject, grade, and term.

Table 2.7. Example of Two-by-Two Classification Table

		True "At-Risk" Status	
		Students Actually "At-Risk"	**Students Actually "Not At-Risk"**
Predicted "At-Risk" Status	**Students Classified as "At-Risk"**	True Positive (TP)	False Positive (FP)
	Students Classified as "Not At-Risk"	False Negative (FN)	True Negative (TN)

Table 2.8. Description of Classification Accuracy Summary Statistics

Statistic	Description*	Interpretation
Overall Classification Accuracy Rate	(TP + TN) / (total sample size)	Proportion of the study sample whose classification on the state test was consistent with that by the MAP Growth cut scores
False Negative (FN) Rate	FN / (FN + TP)	Proportion of not-at-risk students identified by MAP Growth in those observed as at-risk students on the state test
False Positive (FP) Rate	FP / (FP + TN)	Proportion of at-risk students identified by MAP Growth in those observed as not at-risk students on the state test
Sensitivity	TP / (TP + FN)	Proportion of at-risk students identified by MAP Growth in those observed as such on the state test.
Specificity	TN / (TN + FP)	Proportion of not-at-risk students identified by MAP Growth in those observed as such on the state test.
Area Under the Curve (AUC), including the lower and upper bounds of the 95% confidence interval	Area under the receiver operating characteristics (ROC) curve	How well MAP Growth cut scores separate the study sample into at-risk and not-at-risk categories that match those from the state test cut scores. AUC, including the lower and upper bounds of the 95% confidence level, were obtained from the ROC analysis via SAS PROC LOGISTIC.

*FP = false positives. FN = false negatives. TP = true positives. TN = true negatives.

2.6.2. Results

After conducting the 341 classification accuracy analyses for each subject, grade, and term and evaluating these statistics against the NCII rating rubrics (NCII, 2020a), the results concluded that the candidate MAP Growth cut scores corresponding to the 30th percentile rank based on the national norms performed the best for identifying students in need of intensive intervention, given that the bottom 10% of the students on the criterion measure are assumed as students actually in need of intensive intervention. Thus, the candidate cut scores corresponding to the 30th percentile rank are recommended as the MAP Growth universal screening cut scores to identify students at severe risk and in need of intensive intervention. Only the classification accuracy results for these recommended cut scores (i.e., the cut scores that yielded the best classification accuracy) are provided in this report in Table 2.9 and Table 2.10.

Results for Grades 3–8 suggest that the lower bounds of the AUCs for the recommended cut scores are at least 0.9 for all subjects and terms, and most of the specificities and sensitivities are above 0.8. The classification accuracy statistics for Grades K–2 became slightly worse. For example, the lower bounds of AUCs dropped below 0.9 for Grades K–1. For Grade K, while specificities remained above 0.8 across all terms, sensitivities decreased significantly. Grades K–2 students were included as longitudinal data based on the Grade 3 student cohort from each sample, which contributed to these lower classification accuracy results. For those students, the analyses used state assessment scores when they were in Grade 3 but MAP Growth scores when they were in Grades K–2. That is, these MAP Growth tests were taken approximately 12–36 months earlier than the state tests. In general, the longer the time lapse between the criterion measure and MAP Growth, the worse the classification accuracy indices became.

Table 2.9. Classification Accuracy Results Based on the Recommended MAP Growth Universal Screening Cut Scores—Primary Sample

Grade	Term	Recommended MAP Growth Cut Score	Class. Accuracy*	FP	FN	Sensitivity	Specificity	AUC	AUC (LB)	AUC (UB)
Mathematics										
	Fall	133	0.88	0.08	0.59	0.42	0.92	0.83	0.82	0.84
K	Winter	144	0.83	0.16	0.35	0.65	0.85	0.85	0.83	0.86
	Spring	151	0.88	0.09	0.41	0.59	0.91	0.88	0.87	0.89
	Fall	154	0.84	0.15	0.28	0.72	0.85	0.88	0.87	0.89
1	Winter	164	0.79	0.21	0.17	0.83	0.79	0.90	0.89	0.91
	Spring	169	0.83	0.17	0.18	0.82	0.83	0.91	0.90	0.91
	Fall	168	0.81	0.20	0.14	0.86	0.80	0.90	0.89	0.91
2	Winter	177	0.82	0.19	0.12	0.88	0.81	0.91	0.91	0.92
	Spring	182	0.85	0.15	0.17	0.83	0.86	0.92	0.91	0.93
	Fall	181	0.85	0.16	0.13	0.87	0.84	0.92	0.92	0.93
3	Winter	189	0.83	0.19	0.07	0.93	0.82	0.94	0.94	0.95
	Spring	194	0.85	0.16	0.07	0.93	0.84	0.95	0.95	0.95
	Fall	192	0.87	0.12	0.19	0.81	0.88	0.93	0.92	0.93
4	Winter	198	0.86	0.14	0.14	0.86	0.86	0.93	0.93	0.94
	Spring	202	0.88	0.12	0.11	0.89	0.88	0.95	0.95	0.95
	Fall	201	0.86	0.14	0.16	0.85	0.86	0.93	0.92	0.93
5	Winter	206	0.86	0.14	0.15	0.85	0.86	0.93	0.93	0.94
	Spring	210	0.86	0.15	0.11	0.89	0.85	0.94	0.94	0.95
	Fall	206	0.85	0.15	0.13	0.87	0.85	0.93	0.92	0.93
6	Winter	211	0.80	0.21	0.08	0.92	0.79	0.93	0.93	0.94
	Spring	214	0.82	0.19	0.07	0.93	0.81	0.94	0.94	0.95
	Fall	211	0.84	0.17	0.12	0.88	0.83	0.93	0.92	0.93
7	Winter	215	0.82	0.19	0.09	0.91	0.81	0.93	0.93	0.94
	Spring	217	0.84	0.17	0.10	0.90	0.84	0.94	0.93	0.94
	Fall	215	0.82	0.19	0.14	0.86	0.81	0.91	0.90	0.92
8	Winter	218	0.83	0.18	0.14	0.87	0.83	0.92	0.91	0.92
	Spring	220	0.81	0.20	0.10	0.90	0.80	0.92	0.92	0.93

Grade	Term	Recommended MAP Growth Cut Score	Class. Accuracy*	FP	FN	Sensitivity	Specificity	AUC	AUC (LB)	AUC (UB)
Reading										
K	Fall	130	0.85	0.10	0.67	0.33	0.90	0.77	0.75	0.79
	Winter	140	0.83	0.14	0.49	0.51	0.86	0.80	0.79	0.82
	Spring	147	0.79	0.19	0.36	0.64	0.81	0.82	0.80	0.83
1	Fall	149	0.75	0.25	0.23	0.77	0.75	0.84	0.82	0.85
	Winter	159	0.69	0.33	0.12	0.89	0.67	0.86	0.85	0.87
	Spring	164	0.75	0.27	0.13	0.87	0.74	0.88	0.87	0.89
2	Fall	164	0.76	0.26	0.10	0.91	0.74	0.89	0.88	0.90
	Winter	173	0.79	0.21	0.13	0.87	0.79	0.91	0.90	0.91
	Spring	177	0.85	0.15	0.18	0.82	0.85	0.91	0.91	0.92
3	Fall	178	0.83	0.17	0.14	0.86	0.83	0.92	0.91	0.92
	Winter	185	0.82	0.19	0.12	0.88	0.81	0.92	0.92	0.93
	Spring	189	0.84	0.16	0.12	0.88	0.84	0.93	0.93	0.94
4	Fall	188	0.86	0.13	0.21	0.79	0.87	0.91	0.91	0.92
	Winter	194	0.86	0.14	0.17	0.83	0.86	0.92	0.92	0.92
	Spring	196	0.86	0.14	0.17	0.83	0.87	0.93	0.92	0.93
5	Fall	196	0.87	0.12	0.19	0.81	0.88	0.93	0.92	0.93
	Winter	201	0.85	0.16	0.15	0.85	0.85	0.92	0.92	0.93
	Spring	203	0.86	0.14	0.14	0.86	0.86	0.93	0.93	0.94
6	Fall	202	0.85	0.15	0.18	0.82	0.86	0.92	0.91	0.92
	Winter	205	0.84	0.16	0.13	0.87	0.84	0.93	0.92	0.93
	Spring	207	0.84	0.16	0.13	0.87	0.84	0.93	0.92	0.93
7	Fall	206	0.86	0.14	0.19	0.81	0.86	0.92	0.91	0.92
	Winter	209	0.84	0.16	0.17	0.84	0.84	0.92	0.91	0.92
	Spring	210	0.84	0.16	0.18	0.82	0.85	0.92	0.91	0.92
8	Fall	209	0.86	0.14	0.18	0.82	0.86	0.91	0.91	0.92
	Winter	212	0.86	0.13	0.18	0.82	0.87	0.92	0.92	0.93
	Spring	213	0.85	0.16	0.16	0.84	0.85	0.92	0.91	0.93

*Class. Accuracy = overall classification accuracy rate. FP = false positives. FN = false negatives. AUC = area under the ROC curve. AUC(LB) = lower bound of AUC. AUC(UB) = upper bound of AUC.

Table 2.10. Classification Accuracy Results Based on the Recommended MAP Growth Universal Screening Cut Scores—Secondary Sample

Grade	Term	Recommended MAP Growth Cut Score	Class. Accuracy*	FP	FN	Sensitivity	Specificity	AUC	AUC (LB)	AUC (UB)
Mathematics										
K	Fall	133	0.92	0.02	0.83	0.17	0.98	0.82	0.81	0.84
	Winter	144	0.86	0.12	0.43	0.57	0.89	0.86	0.85	0.87
	Spring	151	0.89	0.09	0.40	0.60	0.91	0.88	0.87	0.89
1	Fall	154	0.85	0.14	0.27	0.73	0.86	0.88	0.88	0.89
	Winter	164	0.82	0.19	0.15	0.85	0.81	0.91	0.90	0.91
	Spring	169	0.85	0.15	0.17	0.84	0.85	0.92	0.92	0.93

Grade	Term	Recommended MAP Growth Cut Score	Class. Accuracy*	FP	FN	Sensitivity	Specificity	AUC	AUC (LB)	AUC (UB)
	Fall	168	0.83	0.17	0.19	0.81	0.83	0.90	0.89	0.91
2	Winter	177	0.85	0.15	0.17	0.83	0.85	0.92	0.92	0.93
	Spring	182	0.89	0.11	0.18	0.82	0.89	0.94	0.93	0.94
	Fall	181	0.85	0.16	0.09	0.91	0.84	0.94	0.94	0.95
3	Winter	189	0.86	0.15	0.05	0.95	0.85	0.96	0.96	0.96
	Spring	194	0.88	0.13	0.06	0.95	0.87	0.97	0.96	0.97
	Fall	192	0.89	0.11	0.15	0.85	0.89	0.95	0.95	0.95
4	Winter	198	0.89	0.11	0.10	0.90	0.89	0.96	0.96	0.96
	Spring	202	0.89	0.12	0.06	0.94	0.88	0.97	0.96	0.97
	Fall	201	0.86	0.14	0.10	0.90	0.86	0.95	0.95	0.95
5	Winter	206	0.87	0.14	0.08	0.92	0.87	0.96	0.95	0.96
	Spring	210	0.86	0.15	0.04	0.96	0.85	0.96	0.96	0.97
	Fall	206	0.89	0.11	0.11	0.89	0.89	0.95	0.95	0.96
6	Winter	211	0.86	0.15	0.07	0.93	0.85	0.96	0.96	0.96
	Spring	214	0.87	0.14	0.05	0.95	0.86	0.97	0.97	0.97
	Fall	211	0.89	0.11	0.13	0.87	0.89	0.95	0.95	0.95
7	Winter	215	0.87	0.13	0.10	0.90	0.87	0.95	0.95	0.96
	Spring	217	0.89	0.12	0.10	0.90	0.89	0.96	0.96	0.96
	Fall	215	0.89	0.10	0.20	0.80	0.90	0.94	0.93	0.94
8	Winter	218	0.89	0.11	0.19	0.81	0.90	0.94	0.94	0.94
	Spring	220	0.89	0.11	0.15	0.85	0.89	0.95	0.94	0.95
Reading										
	Fall	130	0.86	0.11	0.63	0.37	0.89	0.75	0.73	0.76
K	Winter	140	0.79	0.20	0.40	0.60	0.80	0.79	0.78	0.81
	Spring	147	0.80	0.19	0.32	0.68	0.81	0.82	0.81	0.84
	Fall	149	0.76	0.24	0.27	0.73	0.76	0.82	0.81	0.83
1	Winter	159	0.70	0.31	0.14	0.86	0.69	0.85	0.85	0.86
	Spring	164	0.77	0.24	0.19	0.81	0.76	0.87	0.86	0.87
	Fall	164	0.79	0.22	0.18	0.82	0.78	0.88	0.87	0.89
2	Winter	173	0.84	0.16	0.22	0.79	0.84	0.90	0.89	0.90
	Spring	177	0.87	0.12	0.27	0.73	0.88	0.90	0.89	0.91
	Fall	178	0.82	0.19	0.17	0.83	0.81	0.90	0.89	0.90
3	Winter	185	0.83	0.17	0.16	0.84	0.83	0.91	0.90	0.91
	Spring	189	0.85	0.15	0.17	0.83	0.85	0.92	0.91	0.92
	Fall	188	0.86	0.13	0.22	0.78	0.87	0.91	0.91	0.92
4	Winter	194	0.86	0.14	0.18	0.82	0.86	0.92	0.92	0.93
	Spring	196	0.86	0.14	0.15	0.85	0.86	0.93	0.92	0.93
	Fall	196	0.85	0.15	0.18	0.82	0.85	0.92	0.91	0.92
5	Winter	201	0.82	0.18	0.13	0.87	0.82	0.92	0.92	0.93
	Spring	203	0.83	0.18	0.12	0.88	0.82	0.93	0.92	0.93

Grade	Term	Recommended MAP Growth Cut Score	Class. Accuracy*	FP	FN	Sensitivity	Specificity	AUC	AUC (LB)	AUC (UB)
	Fall	202	0.84	0.16	0.19	0.81	0.84	0.91	0.91	0.92
6	Winter	205	0.83	0.17	0.17	0.83	0.83	0.91	0.91	0.92
	Spring	207	0.83	0.17	0.16	0.84	0.83	0.92	0.91	0.92
	Fall	206	0.86	0.14	0.17	0.83	0.86	0.92	0.92	0.93
7	Winter	209	0.83	0.18	0.13	0.87	0.83	0.92	0.92	0.92
	Spring	210	0.85	0.15	0.13	0.87	0.85	0.93	0.93	0.94
	Fall	209	0.85	0.15	0.16	0.84	0.85	0.92	0.92	0.93
8	Winter	212	0.85	0.15	0.16	0.84	0.85	0.92	0.92	0.93
	Spring	213	0.85	0.15	0.15	0.85	0.85	0.93	0.92	0.93

*Class. Accuracy = overall classification accuracy rate. FP = false positives. FN = false negatives. AUC = area under the ROC curve. AUC (LB) = lower bound of AUC. AUC (UB) = upper bound of AUC.

3. Universal Screening Cut Scores for Spanish MAP Growth Reading

3.1. Spanish MAP Growth Reading Overview

Piloted in the 2018–2019 school year, the Spanish MAP Growth Reading test was officially released in Fall 2019 covering Grades K–8. It was designed to be a parallel assessment to the English version but also considers aspects of reading that are specific to the Spanish language. Spanish MAP Growth Reading has its own scale that is linked to the existing English scale. Like its English counterpart, Spanish MAP Growth Reading tests are adaptive. Reporting features are also similar. Educators can receive data from both the English and Spanish MAP Growth Reading growth measures if students take both assessments, allowing them to make informed decisions to support their students' learning in both languages.[3] A norming study completed in July 2020 produced student achievement status (i.e., fall, winter, and spring) and growth norms (i.e., fall-to-winter, winter-to-spring, and fall-to-spring within a school year).

3.2. Universal Screening Cut Scores

The Spanish universal screening cut scores were established by linking the English and Spanish MAP Growth Reading test scores and finding the score on the Spanish assessment that corresponded to the recommended universal screening cut scores from the English MAP Growth Reading assessment described in Section 2 of this report. First, a linking study with a small group of Grade 3 students who took both the Spanish MAP Growth Reading pilot test and the English Reading test in Spring 2019 established a connection between scores on the two assessments using the equipercentile procedure. The recommended English universal screening cut scores were then applied to the Spanish assessment to obtain the cut scores. Specifically, the Spanish reading score corresponding to the English Grade 3 spring cut score was at the 40th percentile of the Spanish MAP Growth Reading norms. Using this percentile, cut scores for the other grades and terms were identified.

Table 3.1 presents the Spanish MAP Growth Reading universal screening cut scores that correspond to the 40th percentile of the Spanish MAP Growth Reading norms. Students with Spanish reading scores lower than these cut scores are likely at risk for reading deficiency and in need of intensive intervention.

Table 3.1. Spanish MAP Growth Reading Cut Scores for Universal Screening

	Fall		Winter		Spring	
Grade	Cut Score	Percentile	Cut Score	Percentile	Cut Score	Percentile
K	130	40	140	40	148	40
1	145	40	155	40	163	40
2	165	40	173	40	180	40
3	179	40	185	40	186	40
4	187	40	192	40	195	40
5	194	40	198	40	201	40
6	200	40	202	40	207	40
7	204	40	206	40	211	40
8	207	40	209	40	213	40

[3] Technical details for Spanish MAP Growth Reading are provided in the technical report (NWEA, 2020c).

4. Conclusion

The recommended cut scores to identify students in need of intensive intervention correspond to the 30th percentile on the English MAP Growth Mathematics and Reading assessments and the 40th percentile for Spanish MAP Growth Reading. Students scoring below these cut scores are at risk for severe learning difficulties in a subject. The English MAP Growth cut scores were selected based on classification accuracy analyses that used multiple state assessments as criterion measures. Results were evaluated against the NCII classification accuracy criteria regarding what is considered an effective universal screener (NCII, 2020a), and the best cut scores were selected. The cross-validation results based on the secondary sample were consistent with those from the primary sample, providing evidence that the recommended universal screening cut scores are valid.

Universal screening is paramount in identifying students at risk for academic difficulty in an RTI model. Assessment plays a crucial role and represents the first step to identify students at risk for learning difficulties. For a universal screener to be effective, aside from its technical adequacy, it is imperative to establish benchmarks through a scientifically designed and evidenced-based process. As correct identification of at-risk students is critical to ensure that students receive appropriate tiered intervention, an effective universal screener should “yield a high percentage of true positives while identifying a manageable risk pool by limiting false positives” (Fuchs et al., 2007, p. 312). High sensitivity and specificity of a universal screener will increase the likelihood of true positives and decrease the likelihood of false positives. As demonstrated by the classification accuracy results in this study, the recommended cut scores for both subjects resulted in sensitivity, specificity, and lower bound of the AUC of at least 0.8 for most grades and terms—the highest level of the evaluation criteria described in the NCII rating rubrics (NCII, 2020a).

As a type of prediction, no universal screener can be free from errors. False positives and false negatives tend to occur when screening students, particularly in the early grades where the errors tend to be higher than those for more advanced grades. For example, the classification accuracy statistics are lower for Grades K–2 compared to Grades 3–8, although this can be attributed to the longer time lapse between the criterion measure and MAP Growth for Grades K–2 (i.e., students typically do not start taking the state summative assessment until Grade 3, so the MAP Growth scores for Grades K–2 are from 12–36 months prior). Therefore, it is highly recommended for schools and educators to incorporate other factors in conjunction with student performance against the established benchmarks to identify students in need of intensive intervention (e.g., behavior screening, teacher inputs on students, core course completion/failure, school dropout, GPA). NWEA is committed to conducting more research studies with different criterion measures, particularly for Grades K–2, to determine to what degree the precision of the recommended benchmarks in identifying students in need of intensive intervention holds across different screening approaches. NWEA also plans to collect more data in the near future to update the universal screening cut scores for Spanish MAP Growth Reading.

5. References

Crawford, L. (2014). The role of assessment in a response to intervention model. *Preventing School Failure: Alternative Education for Children and Youth, 58*(4), 230–236.

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children, 73*(3), 311–330.

National Center on Intensive Intervention (NCII). (n.d.). *Intensive intervention & multi-tiered system of supports (MTSS)*. Retrieved October 1, 2020, from https://intensiveintervention.org/intensive-intervention/multi-tiered-systems-support

National Center on Intensive Intervention (NCII). (2020a). *Academic screening tools chart rating rubrics*.

National Center on Intensive Intervention (NCII). (2020b). *2020 call for submissions of academic screening tools*. American Institutes for Research (AIR). https://intensiveintervention.org/sites/default/files/NCII_AcadScreen_CallForSubmissions_2020-06-30.pdf

NWEA. (2020a). *MAP Growth technical report*. NWEA Research Report.

NWEA. (2020b). Linking study report: *Predicting performance on the Indiana learning evaluation readiness network (ILEARN) based on MAP Growth scores*. NWEA Research Report.

NWEA. (2020c). *Spanish MAP Growth reading technical report*. NWEA Research Report.

Reardon, S. F., Kalogrides, D., & Ho, A. (2017). *Linking U.S. school district test score distributions to a common scale* (CEPA working paper No.16-09). Stanford Center for Education Policy Analysis. https://cepa.stanford.edu/sites/default/files/wp16-09-v201706.pdf

Thum, Y., & Kuhfeld, M. (2020). *NWEA 2020 MAP Growth achievement status and growth norms for students and schools*. NWEA Research Report.

Ysseldyke, J., Burns, M. K., Scholin, S. E., & Parker, D. C. (2010). Instructionally valid assessment with response to intervention. *TEACHING Exceptional Children*, *42*, 54-61.