Curriculum Associates RESEARCH

i-Ready Personalized Instruction and State Assessment Performance

Reading and Mathematics

Ethan Young, Ph.D. and Kelsey Young, Ph.D.

Research Report, December 2025

i-Ready Personalized Instruction and State Assessment Performance

Introduction

State assessment performance is a key metric in accountability systems. As such, state and district leaders often require evidence of how instructional products may support this performance. Instructional products like the *i-Ready Diagnostic* and *i-Ready Personalized Instruction* (i.e., *i-Ready* PI), however, are state-agnostic and designed to serve broader populations.

To better identify how these tools aid state assessment performance, Curriculum Associates (CA) conducts complimentary lines of state-specific research. Comprehensive linking studies evaluate the correlation between the *i-Ready Diagnostic* and state assessments. These studies provide foundational evidence that the *i-Ready Diagnostic* is a valid measure of reading and mathematics, strongly overlaps with state assessments, and can be used as an indicator of student performance on state assessments.

State Impact Research (SIR) studies quantify the association between *i-Ready* PI usage and state summative exams. They are designed to evaluate whether student engagement with *i-Ready* PI (i.e., fidelity use) is associated with higher scores on state exams. Across states, grades, and subjects, students using *i-Ready* PI with fidelity—for at least 30 minutes a week, with a 70 percent pass rate, and for at least 18 weeks across the school year—tend to score higher on state summative exams in reading and mathematics than their peers.

This existing work largely focuses on states in isolation, answering questions about *i-Ready* PI and state assessments on a state-by-state basis. However, pooling data across states allows for evaluations on a broader scale, across a wider range of students, and presents new opportunities. For example, where individual state samples may be under-powered, collating data across all available state data creates samples large enough to ask new questions about "for whom, and under what conditions" *i-Ready* PI fidelity use is associated with state assessment performance.

Despite these opportunities, there are also clear challenges. Each state designs and validates their own summative assessments for reading and mathematics. This makes exam scores across states incomparable, even when they measure the same construct (e.g., reading and mathematics). To address these challenges, CA leveraged psychometric techniques to better understand the association between *i-Ready* PI and state assessment performance across all SIR studies. Specifically, student data were pooled across states and state assessment scores transformed using equipercentile linking. This linking method places students' mathematics and reading scores on a common metric, making cross-state comparisons possible. This allows us to examine the association between *i-Ready* PI and state performance across a wide range of states and students. It also allows for a close examination of how *i-Ready* PI operates among key student groups including English Learners, students with disabilities, students who are economically disadvantaged, or from different racial and ethnic groups.

Curriculum Associates RESEARCH

Data and Measures

The data for the current research leveraged three sources of information: fall *i-Ready Diagnostic* scores, *i-Ready* PI usage information, and state assessment data, including performance and student demographics. For the current research, we combined 25 state assessment datasets from the 2021–2022, 2022–2023, and 2023–2024 academic years with *i-Ready* PI usage data from the corresponding year (see Table 1).

State Assessment Variables

State-specific variables included state test scores obtained from state testing in the spring for both reading and mathematics, as well as student-level demographic information, where available. State scores are test scores on state assessments reported on state-specific scales. Demographic information includes race/ethnicity, English Learner status, disability status, and economically disadvantaged status.

i-Ready Diagnostic-specific variables included fall Diagnostic scores for Reading and for Mathematics. The *i-Ready Diagnostic* was developed to measure student knowledge, identify student skill level across different domains, and provide a mechanism for delivering instructional guidance (Curriculum Associates, 2018). Student results are reported as scale scores, placement levels, and norm-referenced percentile scores. *i-Ready Diagnostic* scale scores are linear transformations of logit values. For each assessment in Reading and in Mathematics, an overall score is calculated, as are domain scores for each content strand. Scale scores can range in value from 100 to 800.

i-Ready Personalized Instruction

To measure *i-Ready* PI use, we created a set of summary variables to aggregate one academic year of use. To identify students who used as recommended, we created a fidelity usage variable. Specifically, we defined recommended use as meeting three criteria: 1) completed *i-Ready* PI lessons with at least a 70 percent pass rate, 2) used *i-Ready* PI for 18 or more weeks across the academic year, and 3) used *i-Ready* PI for at least 30 minutes per week. Students who met all criteria were coded as fidelity users (1), whereas all other students were coded as non-fidelity users (0). For the purposes of this research, our main analysis utilizes this definition of fidelity use.

Data Preparation

To create the analytic sample, we matched students in state assessment datasets to their corresponding *i-Ready* PI usage data by student, subject (i.e., reading or mathematics), and academic year (i.e., 2021–2022, 2022–2023, or 2023–2024). Then, we inspected usage summary variables for outliers. To adjust for very low or very high use, we only included students whose use was within 1.5 times the interquartile range of each usage variable. Because *i-Ready* PI use varies by state, subject, and student grade level, we applied our outlier criteria to students within each grade, subject, and state combination.

Our final set of state samples included students who had fall *i-Ready Diagnostic* scores, state test data, and *i-Ready* PI usage variables within grade, subject, and state-specific ranges (within 1.5 times the interquartile range) of each usage variable. Table 1 shows sample sizes for each state and subject combination.

Curriculum Associates RESEARCH

Table 1. Aggregated State Assessment Sample by Study Year and State for Reading and Mathematics

| State	Study Year	Reading			Mathematics		
		Students	Schools	Districts	Students	Schools	Districts
Alaska	2022–2023	7,974	86	4	22,821	129	4
Arizona	2021–2022	10,572	68	11	14,485	80	12
Colorado	2021–2022	33,925	186	13	33,201	187	13
Florida	2023–2024	33,856	121	14	28,607	115	12
Georgia	2021–2022	58,654	192	9	64,806	222	10
Georgia	2023–2024	70,047	239	12	67,783	233	11
Illinois	2022–2023	10,019	61	9	13,945	72	10
Iowa	2022–2023	10,547	52	7	14,180	64	11
Kentucky	2021–2022	16,735	81	11	16,445	81	11
Louisiana	2021–2022	17,031	108	17	33,018	147	20
Maryland	2023–2024	26,600	207	5	18,080	143	7
Massachusetts	2021–2022	7,884	45	12	8,094	45	12
Michigan	2022–2023	25,843	137	11	26,732	136	11
Missouri	2022–2023	23,074	130	17	26,060	135	19
New York	2022–2023	23,333	125	22	29,184	154	26
North Carolina	2021–2022	22,432	102	11	32,440	140	11
Ohio	2022–2023	43,517	181	14	42,780	181	14
Ohio	2023–2024	15,369	96	2	15,383	96	2
Pennsylvania	2023–2024	24,996	125	10	33,500	149	10
SBAC	2021–2022	145,038	621	53	156,010	668	61
South Carolina	2022–2023	82,703	295	34	92,633	304	36
Tennessee	2022–2023	21,427	183	10	23,844	179	10
Texas	2022–2023	18,421	101	14	33,181	157	16
Texas	2023–2024	50,476	338	11	68,158	355	12
Virginia	2021–2022	785	5	2	10,903	53	5
West Virginia	2022–2023	31,021	187	22	30,502	185	23
Wisconsin	2023–2024	2,773	26	11	8,007	42	16

Placing State Assessments on a Common Metric

To make state assessment scores comparable across states, CA leveraged equipercentile linking (Kolen & Brennan, 2004). This technique can be used to convert scores on one test to the associated score on another test. It does so by finding scores on each test that have the same percentile rank. This can be used to generate a crosswalk table of scores from one test to another. Equipercentile linking is a field gold standard for linking tests scores when tests do not share the same test blueprint or items. It can handle non-linear relationships and is generally a robust method of linking.

In the current work, we leveraged findings from existing linking studies, which contain samples of students who took both the *i-Ready Diagnostic* and a state exam within the same testing window. Student scores on each test were converted to percentile ranks and used to find scores on each test with the same percentile rank. This procedure produces a crosswalk of all possible *i-Ready Diagnostic* scores and their corresponding score on the state exam. These tables allow us to

convert all state scores in the pooled SIR dataset to a common metric—the *i-Ready Diagnostic* scale (see Figure 1).

Figure 1. Example Transformation of State Assessment Scales to a Common Metric
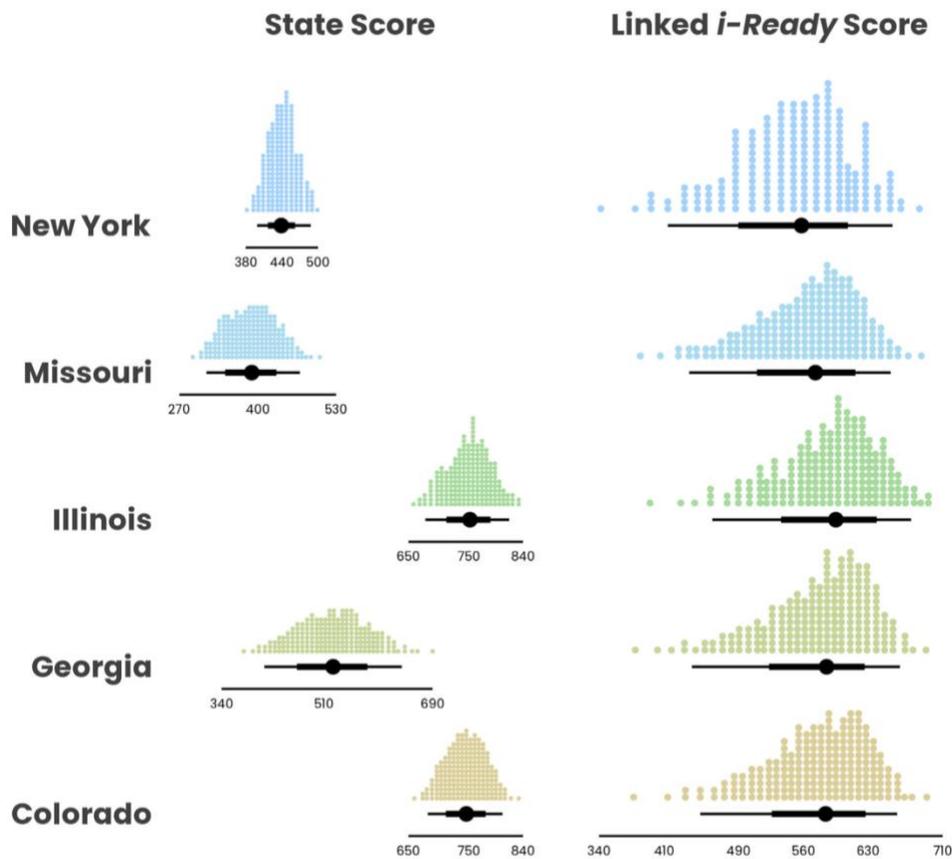
Note: Example uses Grade 5 reading scores

Analysis

After pooling all state data and converting state assessment scores to the *i-Ready Diagnostic* scale scores, the primary analytic goal was to test the relation between *i-Ready* PI and state assessment scores across all states. The analytic approach was nearly identical to individual state analyses with one exception. Specifically, as with individual state analyses, models included the fall *i-Ready Diagnostic* scores to account for prior student performance, and each grade level and subject was analyzed individually. However, given the larger sample sizes and increased number of schools included, data were modeled with a different technique to account for potential school-level confounds. For example, schools with high rates of *i-Ready* PI fidelity use might also implement other programs well and/or have more support and resources. Thus, the effect of student *i-Ready* PI fidelity may be conflated with a supportive school context and drive higher state assessment scores.

Figure 2. Conceptual Model for Understanding *i-Ready* PI Use and State Test Scores
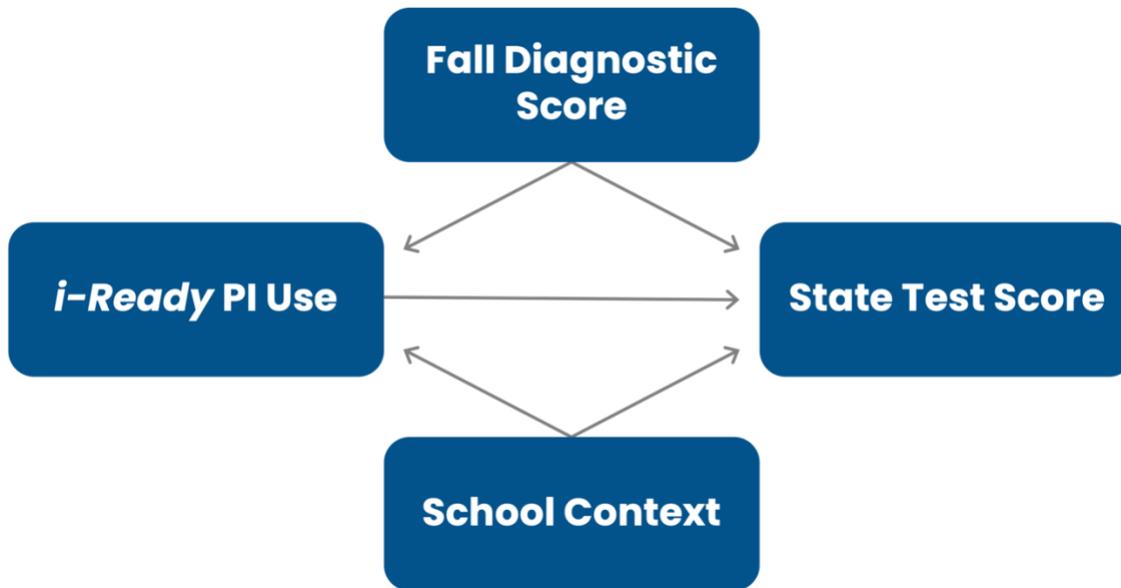
To address this issue, data were modeled using a full-fixed effects approach. This approach is an alternative to mixed modeling approaches that handle nested data structures common in education research (e.g., students nesting in schools). Whereas both modeling approaches handle nested data structures, full fixed effects models use group mean centering to remove between school variation. That is, all student-level variables are adjusted for their grade, school, state, and academic year group means. In other words, all student variables are expressed as deviation scores from their school's average score. For example, a student with an adjusted state assessment score = 0 would have scored exactly average in their school. Negative values reflect scores lower than school-level averages, whereas positive values reflect higher scores than the school average. In effect, this procedure removes between school variation and focuses on estimating *within* school variation.

There are several advantages to full fixed effects models. First, compared with mixed models that included fixed and random effects for nesting variables, they are computationally very efficient. Second, they convert model estimates in a way that aids interpretation; the association between *i-Ready* PI in this type of model tests whether student state assessment scores tend to increase compared to the school average. Finally, by removing between school variation, school-level confounds that could bias the model are controlled for, including possible unmeasured school-related variables.

To explore these associations among subgroups, this analysis was repeated for across a set of student populations including English Learners, students with disabilities, economically disadvantaged students, and students in different racial/ethnic groups. Due to missing data and small subsamples, we include analyses for the following racial/ethnic groups: Asian, Black, Hispanic, Two or More Races, and White (see Table 2). These analyses were identical to the focal analysis with the addition of an interaction term to estimate the association between fidelity usage and subgroup status.

Curriculum Associates RESEARCH

Table 2. Sample Size and Percent Missing by Subgroup

Reading

	Subgroup		Grade 3 (n = 160,831)		Grade 4 (n = 162,154)		Grade 5 (n = 157,265)		Grade 6 (n = 131,759)		Grade 7 (n = 100,452)		Grade 8 (n = 80,620)	
			n	%	n	%	n	%	n	%	n	%	n	%
	Disability Status	No	115,619	71.9%	116,649	71.9%	112,626	71.6%	95,219	72.3%	72,952	72.6%	58,401	72.4%
		Yes	21,004	13.1%	21,573	13.3%	20,938	13.3%	16,596	12.6%	11,940	11.9%	9,169	11.4%
		Missing	24,208	15.1%	23,932	14.8%	23,701	15.1%	19,944	15.1%	15,560	15.5%	13,050	16.2%
	Economically Disadvantaged Status	No	50,565	31.4%	49,919	30.8%	48,008	30.5%	40,546	30.8%	30,115	30.0%	23,905	29.7%
		Yes	74,277	46.2%	74,227	45.8%	71,921	45.7%	60,033	45.6%	41,650	41.5%	32,829	40.7%
		Missing	35,989	22.4%	38,008	23.4%	37,336	23.7%	31,180	23.7%	28,687	28.6%	23,886	29.6%
	English Learner Status	No	114,018	70.9%	115,179	71.0%	112,751	71.7%	95,831	72.7%	74,037	73.7%	60,042	74.5%
		Yes	21,512	13.4%	21,572	13.3%	19,484	12.4%	16,332	12.4%	9,634	9.6%	5,939	7.4%
		Missing	25,301	15.7%	25,403	15.7%	25,030	15.9%	19,596	14.9%	16,781	16.7%	14,639	18.2%
Reading	Racial/Ethnic Group	American Indian or Alaska Native	800	0.5%	702	0.4%	680	0.4%	556	0.4%	317	0.3%	264	0.3%
		Asian	8,493	5.3%	8,277	5.1%	7,927	5.0%	6,652	5.0%	5,023	5.0%	4,084	5.1%
		Black	32,239	20.0%	32,887	20.3%	32,546	20.7%	2,7341	20.8%	21,903	21.8%	16,915	21.0%
		Hispanic	45,681	28.4%	47,024	29.0%	45,864	29.2%	39,493	30.0%	26,898	26.8%	21,832	27.1%
		Native Hawaiian or Pacific Islander	567	0.4%	519	0.3%	491	0.3%	339	0.3%	211	0.2%	146	0.2%
		Two or More Races	7,913	4.9%	7,786	4.8%	7,179	4.6%	57,00	4.3%	4,516	4.5%	3,650	4.5%
		White	59,613	37.1%	59,013	36.4%	57,066	36.3%	46,231	35.1%	36,803	36.6%	29,878	37.1%
		Missing	5,525	3.4%	5,946	3.7%	5,512	3.5%	5,447	4.1%	4,781	4.8%	3,851	4.8%

Mathematics

|	Subgroup		(n = 194,736)		(n = 194,844)		(n = 189,124)		(n = 147,333)		(n = 114,513)		(n = 90,738)	
	Disability Status	No	140,350	72.1%	141,280	72.5%	137,262	72.6%	108,610	73.7%	83,775	73.2%	67,249	74.1%
		Yes	26,048	13.4%	26,772	13.7%	26,091	13.8%	18,647	12.7%	14,114	12.3%	11,252	12.4%
		Missing	28,338	14.6%	26,792	13.8%	25,771	13.6%	20,076	13.6%	16,624	14.5%	12,237	13.5%
	Economically Disadvantaged Status	No	62,022	31.8%	61,861	31.7%	60,795	32.1%	46,972	31.9%	36,282	31.7%	29,194	32.2%
		Yes	91,991	47.2%	91,682	47.1%	88,794	47.0%	69,308	47.0%	50,450	44.1%	41,247	45.5%
		Missing	40,723	20.9%	41,301	21.2%	39,535	20.9%	31,053	21.1%	27,781	24.3%	20,297	22.4%
	English Learner Status	No	135,037	69.3%	137,252	70.4%	134,655	71.2%	108,393	73.6%	86,405	75.5%	69,691	76.8%
Mathematics		Yes	31,565	16.2%	31,166	16.0%	28,353	15.0%	18,483	12.5%	10,392	9.1%	7,803	8.6%
		Missing	28,134	14.4%	26,426	13.6%	26,116	13.8%	20,457	13.9%	17,716	15.5%	13,244	14.6%
	Racial/Ethnic Group	American Indian or Alaska Native	1,125	0.6%	1,087	0.6%	1,125	0.6%	891	0.6%	677	0.6%	487	0.5%
		Asian	9,685	5.0%	9,506	4.9%	9,112	4.8%	7,038	4.8%	5,019	4.4%	3,431	3.8%
		Black	36,940	19.0%	36,941	19.0%	36,322	19.2%	30,790	20.9%	26,132	22.8%	20,224	22.3%
		Hispanic	58,827	30.2%	59,631	30.6%	57,828	30.6%	41,548	28.2%	27,116	23.7%	23,104	25.5%
		Native Hawaiian or Pacific Islander	701	0.4%	715	0.4%	697	0.4%	506	0.3%	345	0.3%	303	0.3%
		Two or More Races	9,386	4.8%	9,414	4.8%	8,746	4.6%	6,720	4.6%	5,326	4.7%	4,272	4.7%
		White	72,394	37.2%	71,390	36.6%	69,579	36.8%	5,4387	36.9%	44,942	39.2%	34,561	38.1%
		Missing	5,678	2.9%	6,160	3.2%	5,715	3.0%	5,453	3.7%	49,56	4.3%	4,356	4.8%

Results

Overall Differences

Across grades and subjects, analyses revealed a consistent, positive association between use of *i-Ready* PI with fidelity and state assessments. Standardized differences between fidelity and non-

fidelity student groups ranged between .14 to .24, indicating students using *i-Ready* PI with fidelity tended to score higher on state assessments than students using less consistently, even after accounting for prior achievement and school context (see Figure 3). In education research, group differences of this magnitude are considered medium to large (Kraft, 2020).

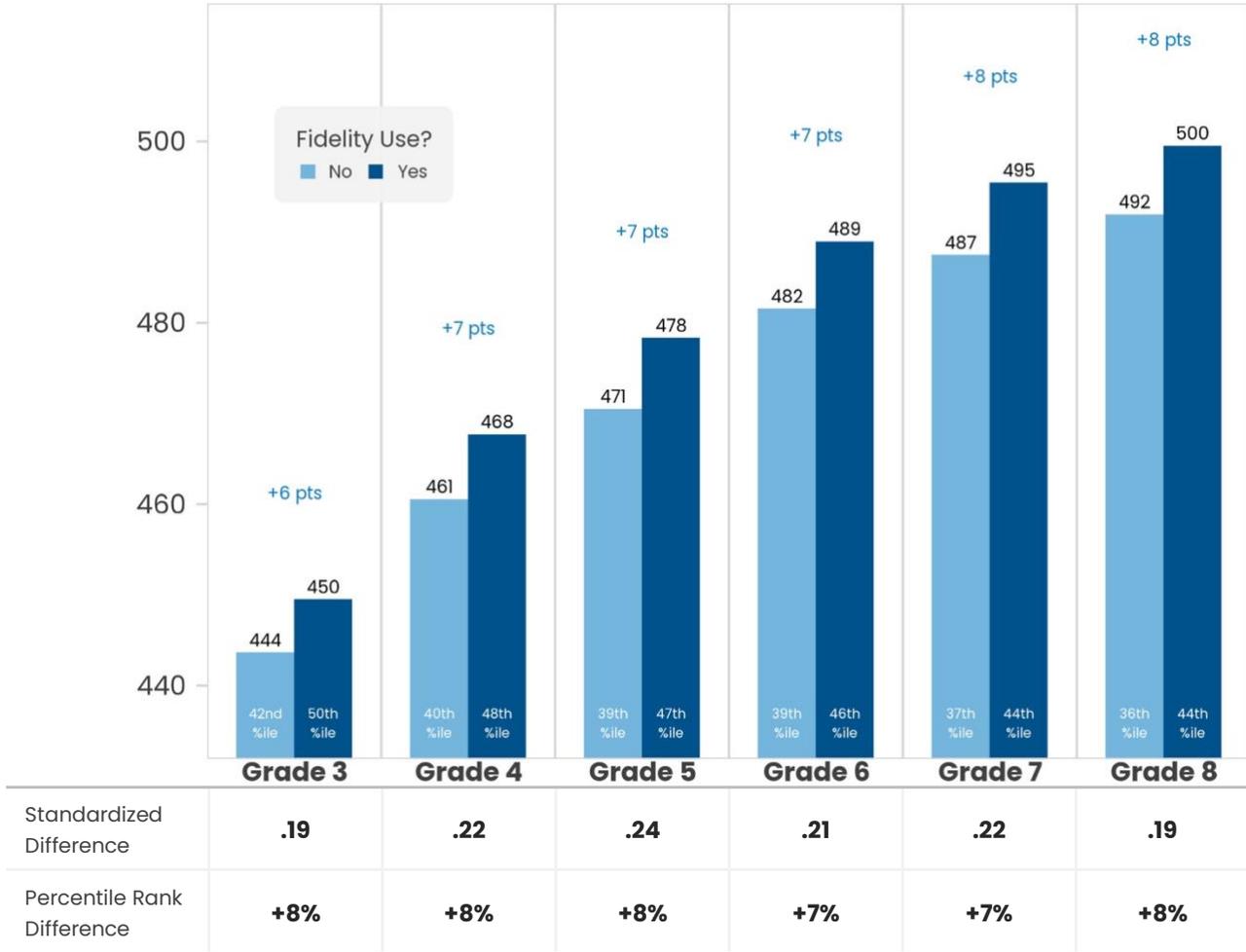Figure 3. Standardized Score Differences between Non-Fidelity and Fidelity *i-Ready* PI Use
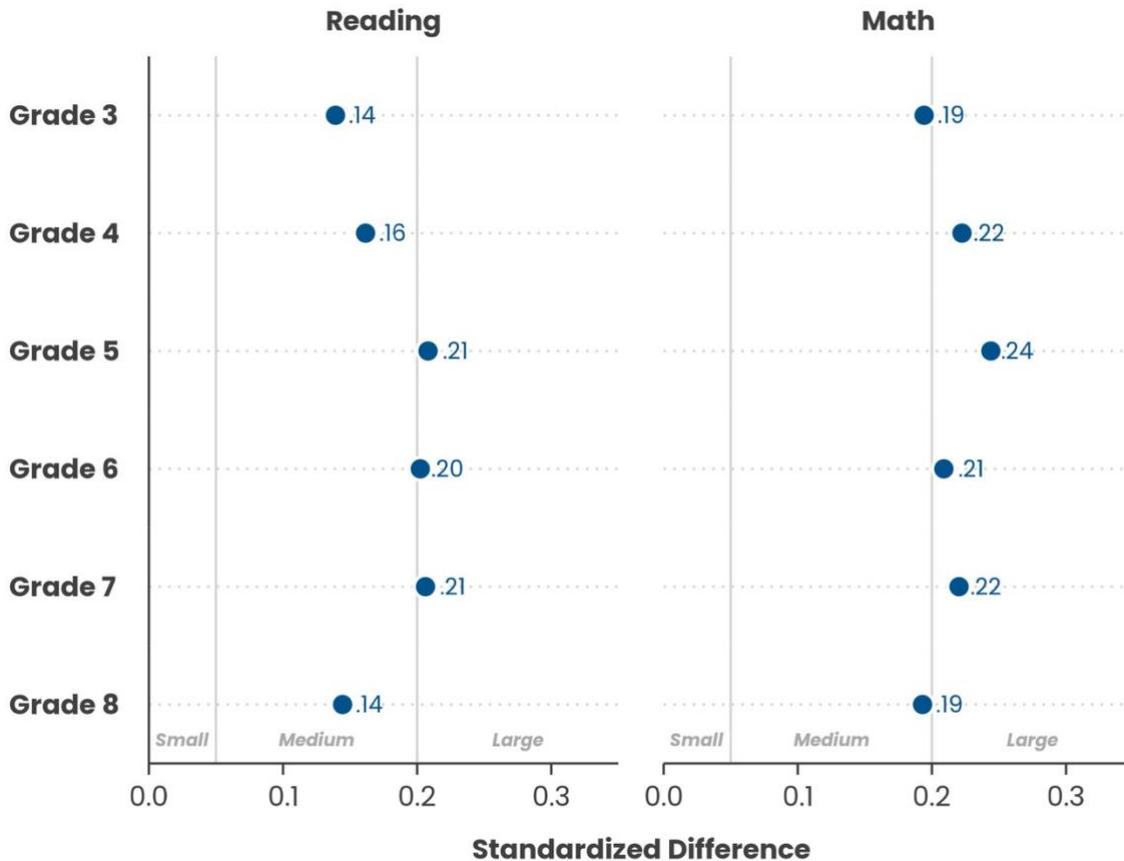
To aid interpretation, we examined the state score estimates expressed as *i-Ready Diagnostic* scores and translated them to percentiles. Students using with fidelity scored anywhere from six to 11 points higher on the *i-Ready* scale than their peers using less consistently. Expressed as percentiles, fidelity users tended to score anywhere from four to eight percentile points higher than their peers (see Figures 4 and 5).

Figure 4. Translated Reading State Scores *by i-Ready* PI Use
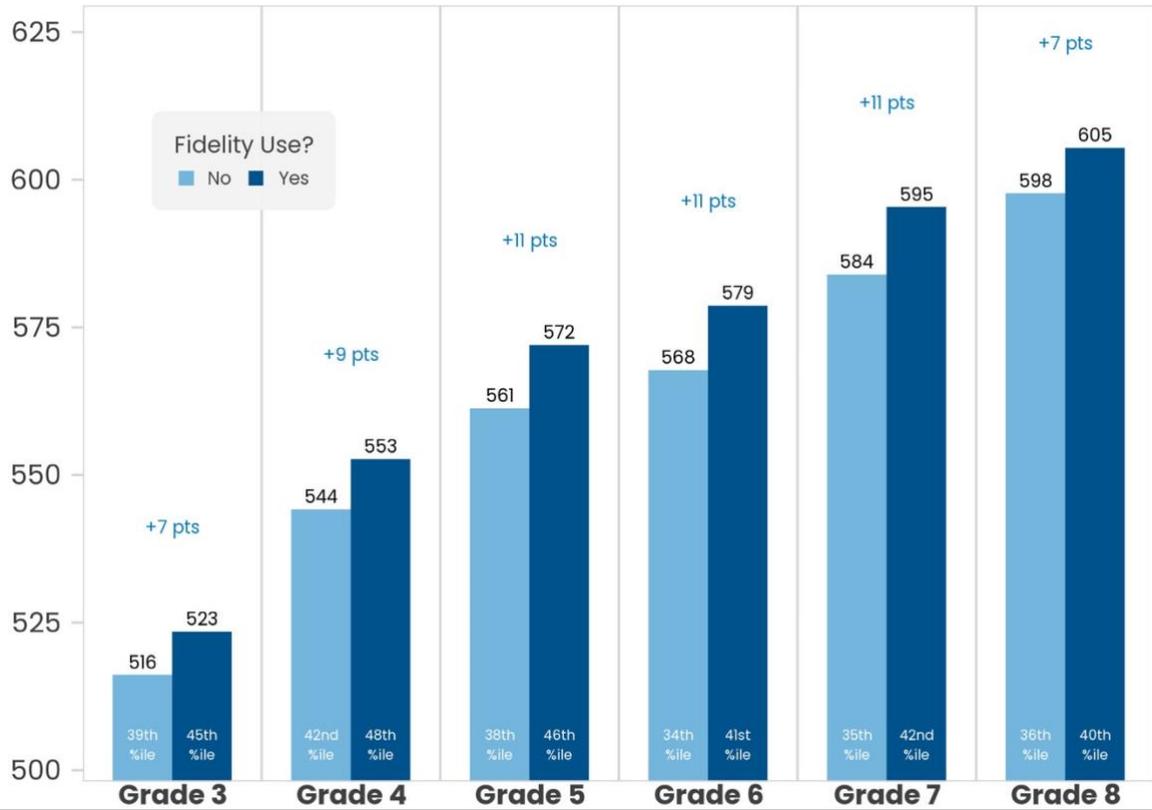
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Standardized Difference	.14	.16	.21	.20	.21	.14
Percentile Rank Difference	+6%	+6%	+8%	+7%	+7%	+4%

Figure 5. Translated Mathematics State Scores by *i-Ready* PI Use

|	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Standardized Difference	.19	.22	.24	.21	.22	.19
Percentile Rank Difference	+8%	+8%	+8%	+7%	+7%	+8%

Differences by Key Student Groups

We also evaluated the association between fidelity use and state assessment performance among key student populations, including racial/ethnic groups (i.e., Asian, Black, Hispanic, Two or More Races, White), students with disabilities, English Learner students, and students who are economically disadvantaged. We conducted one analysis per subgroup type and tested how students using with fidelity compared with students who used less consistently within the same group. Findings revealed a consistent, positive association between fidelity use and state assessment performance within student subgroup. Across grades and subjects, standardized differences between fidelity and non-fidelity users were anywhere from .07 to .26, suggesting fidelity users in each group tended to score higher on state exams compared with their peers from the same group. These differences are considered small to large, with most falling in the medium to large range. Expressed as *i-Ready Diagnostic* score differences, fidelity users tended to score between four and 14 points higher than their peers from the same group (see Table 3).

Figure 6. Standardized Score Differences between Non-Fidelity and Fidelity *i-Ready* PI Use by Subpopulation
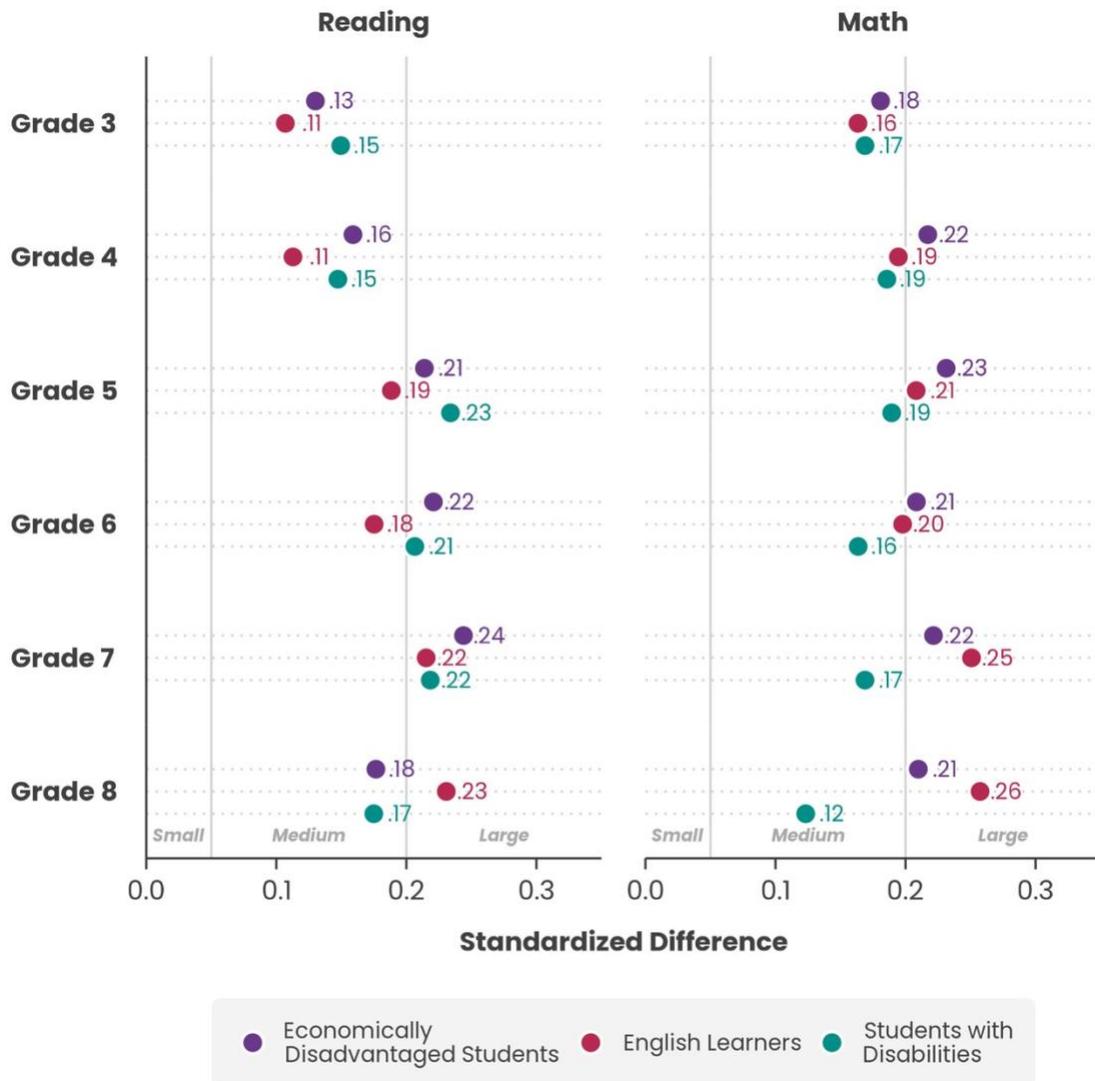
Figure 7. Standardized Score Differences between Non-Fidelity and Fidelity *i-Ready* PI Use by Race/Ethnicity Group
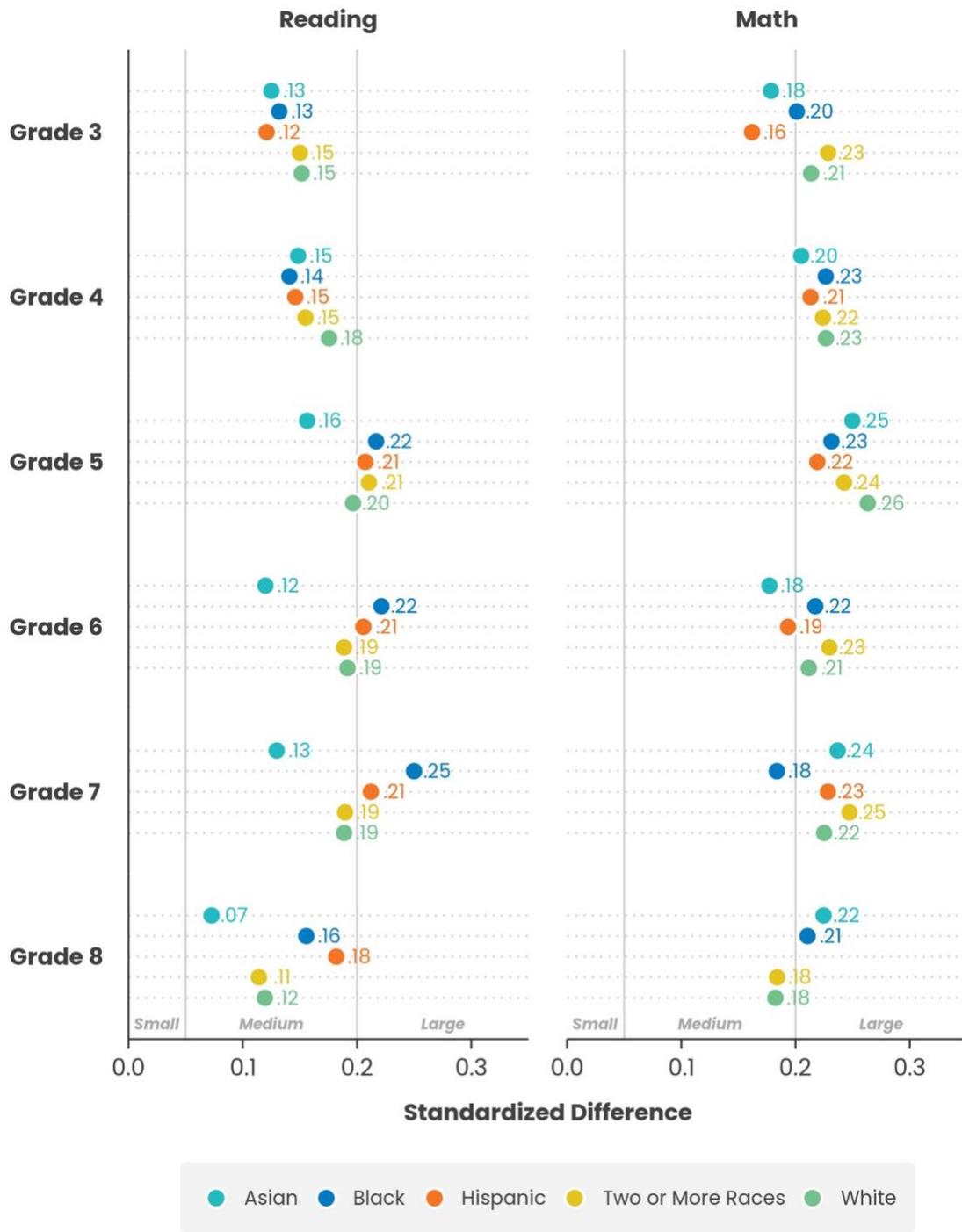
Note: The Grade 8 mathematics Hispanic sample was too small to evaluate state score differences.

Table 3. Translated Reading and Mathematics State Scores by *i-Ready* PI Use and Student Group

Subgroup	Grade	Reading				Mathematics			
		Non-Fidelity	Fidelity	Point Difference	Standardized Difference	Non-Fidelity	Fidelity	Point Difference	Standardized Difference
Economically Disadvantaged Students	3	512	519	7	0.13	440	445	5	0.18
	4	540	548	8	0.16	458	465	7	0.22
	5	556	567	11	0.21	468	475	7	0.23
	6	565	577	12	0.22	480	487	7	0.21
	7	578	592	14	0.24	485	493	8	0.22
	8	595	604	9	0.18	488	496	8	0.21
English Learners	3	514	519	5	0.11	444	449	5	0.16
	4	541	547	6	0.11	460	466	6	0.19
	5	560	570	10	0.19	469	476	7	0.21
	6	563	572	9	0.18	478	485	7	0.20
	7	578	590	12	0.22	486	495	9	0.25
	8	591	603	12	0.23	491	501	10	0.26
Students with Disabilities	3	510	517	7	0.15	440	445	5	0.17
	4	534	542	8	0.15	457	463	6	0.19
	5	550	562	12	0.23	467	473	6	0.19
	6	556	567	11	0.21	478	484	6	0.16
	7	571	583	12	0.22	486	492	6	0.17
	8	582	592	10	0.17	489	494	5	0.12
Asian	3	520	526	6	0.13	448	454	6	0.18
	4	550	557	7	0.15	465	471	6	0.20
	5	568	576	8	0.16	475	483	8	0.25
	6	576	582	6	0.12	486	492	6	0.18
	7	593	600	7	0.13	492	500	8	0.24
	8	606	610	4	0.07	499	508	9	0.22
Black	3	510	517	7	0.13	440	446	6	0.20
	4	538	545	7	0.14	457	464	7	0.23
	5	555	566	11	0.22	467	475	8	0.23
	6	562	574	12	0.22	478	486	8	0.22
	7	578	592	14	0.25	483	490	7	0.18
	8	592	601	9	0.16	490	498	8	0.21
Hispanic	3	515	522	7	0.12	444	449	5	0.16
	4	543	551	8	0.15	460	467	7	0.21
	5	561	572	11	0.21	470	477	7	0.22
	6	567	578	11	0.21	481	487	6	0.19
	7	584	595	11	0.21	486	494	8	0.23
	8	597	607	10	0.18	–	–	–	–
Two or More Races	3	515	523	8	0.15	443	450	7	0.23
	4	543	551	8	0.15	460	467	7	0.22
	5	561	572	11	0.21	470	478	8	0.24
	6	568	578	10	0.19	481	489	8	0.23
	7	585	596	11	0.19	487	496	9	0.25
	8	599	605	6	0.11	494	501	7	0.18
White	3	518	526	8	0.15	445	451	6	0.21
	4	546	555	9	0.18	462	469	7	0.23
	5	563	573	10	0.20	472	480	8	0.26
	6	570	580	10	0.19	484	491	7	0.21
	7	587	598	11	0.19	490	498	8	0.22
	8	600	607	7	0.12	496	503	7	0.18

Discussion

Consistent with prior research, use of *i-Ready* PI with fidelity is associated with improved performance on state assessments. Despite variability in state assessments, including scales, items, and benchmarks for proficiency, pooled results show medium to large standardized score differences between groups (i.e., non-fidelity and fidelity). This pattern held across subpopulations, indicating *i-Ready* PI may be beneficial for many student groups in improving state assessment performance.

Though results are correlational, they suggest fidelity implementation of interventions is important for student academic performance. Schools in which many students use with fidelity likely capture other beneficial academic practices (e.g., appropriately individualized instruction, strong classroom management, thoughtful scheduling). Fidelity implementation can then be viewed as an indicator of strong academic practices associated with improved academic performance while being a measurable goal for which to aim.

Limitations

All results are correlational, testing if *i-Ready* PI use is associated with differences in state scores. Analyses were not designed to test causal effects of *i-Ready* PI on state scores. Though the modeling approach attempts to control for potential school-level confounds, it cannot completely isolate the association between *i-Ready* PI and performance. In addition, though placing all students on a common metric allows for comparison across states, this procedure is not a perfect translation and can introduce error in students' estimated scores. Finally, state assessment samples may not be representative of the state populations on a variety of metrics including but not limited to geographic locale, socioeconomic status, demographic makeup, and/or overall state proficiency levels. See state-specific reports for more information.

References

Curriculum Associates. (2018). *i-Ready assessments technical manual.* Author.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking.* Springer.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798