

Using Existing Data to Create New Diagnostic Measures for States and Districts:

A review of three research-to-practice partnerships that developed measures at the student, educator, school, and system levels

Agenda

1. Introduction to the REL Mid-Atlantic and our research to practice partnerships
2. Using Student-Level Social and Emotional Learning Data to Develop a “Loved, Challenged, and Prepared” Measure
3. Using Kindergarten Entry Assessments to Track City Progress in Promoting Reading Proficiency
4. Measuring School Performance for Early Elementary Grades in Maryland
5. Discussion and questions



Brian Gill
REL Mid-Atlantic
Project Director

REL Mid-Atlantic

- Goals of the REL Program
 - Build partnerships that state and local education agencies in each region to help build capacity to use research evidence to inform education decision making
 - The ultimate goal is to improve student outcomes
- Today's discussion highlights 3 partnerships, in 3 different areas of the region that worked together to use existing data to create new diagnostic measures at the student, educator, school, and system levels



Chris Boccanfuso
Contracting Officer's
Representative
Department of Education

Using Student-Level Social and Emotional Learning Data to Develop a “Loved, Challenged, and Prepared” Measure

A partnership with the District of Columbia Public Schools (DCPS) and REL Mid-Atlantic

Tim Kautz, Mathematica
Kathleen Feeney, Mathematica

The REL Mid-Atlantic Presenters



Tim Kautz, Mathematica



Kathleen Feeney, Mathematica

Project purpose: Why care about social and emotional learning (SEL) competencies?

- They rival academic achievement in predicting success for many life outcomes.¹
- They appear to be malleable in grades K to 12 and can be improved through interventions.²
- They are key outcomes of DCPS's current strategic priority to “educate the whole child.”³



Source: DCPS, 2017.

¹ Heckman & Kautz, 2012. ² Kautz et al., 2014. ³ DCPS, 2017.

Tracking progress toward strategic goals

- DCPS wants to know how it could best use SEL data to do the following:
 - Identify student groups who need additional supports.
 - Track progress in students, schools, and the district as a whole.
 - Evaluate SEL-related practices and programs.
- DCPS administered a new district-wide survey in 2018 to measure SEL competencies.¹

¹ The survey was developed by Panorama Education.

Survey administered to DCPS students in grades 3 to 12

- It covered SEL competencies (for example, perseverance) and SEL supports (for example, “rigorous expectations”).
- It asked students the extent to which they agree to questions such as the following:
“If you fail at an important goal, how likely are you to try again?”

Not at all likely	Slightly likely	Somewhat likely	Quite likely	Extremely likely
------------------------------	----------------------------	----------------------------	-------------------------	-----------------------------

- The response rate was 72 percent (n = 20,030).¹
- Complementary teacher and parent surveys were also administered.

¹ DCPS, 2018b.

Primary goal of the project and partnership

Develop an index of the extent to which students feel “loved, challenged, and prepared” (LCP)

- DCPS partnered with Mathematica through a Regional Educational Laboratory Mid-Atlantic **coaching task** to explore the properties of the district’s survey data.
- Index priorities
 - Useful: DCPS will understand how to use the index to track progress now and in the future.
 - Transparent: DCPS can provide clear information to stakeholders.

Research questions

1. Is the survey reliable and valid for DCPS students?
2. What business rules should we use to construct an index that is transparent and captures DCPS's constructs of LCP?
3. How sensitive is the index to decisions about the rules?

Methods and findings

Step 1: Validate the survey

- Validation helps ensure that the survey items are appropriate for inclusion in summary measures.
- Though Panorama analyzed validity of their surveys, additional analyses helped ensure that the survey is valid for DCPS's students.
 - DCPS modified the original survey.
 - DCPS students might differ from students in previous validation analyses.

Summary of validation exercise

Statistic	Criterion	DCPS's results	Panorama validation
Do the items in each topic capture the same construct?			
Cronbach's alpha	> 0.70 (Bland & Altman, 1997)	> 0.73 for all scales	> 0.68 for all scales
Factor loading for each item	> 0.40 (Stevens, 2012)	> 0.40 for all items	Not reported
Do the groupings of items into topics fit the data well overall?			
Root mean square error of approximation	< 0.05 (Browne & Cudeck, 1992)	0.04	Model "fit well" for all scales
Comparative fit index	> 0.90 (Brown, 2015; Bentler, 1990)	0.92	Model "fit well" for all scales
Tucker-Lewis index	> 0.90 (Brown, 2015; Bentler, 1990)	0.90	Not reported

Step 2: Approach to mapping survey topics to LCP components

LCP alignment	Topic
I feel loved	Sense of belonging (e.g., connections to adults at school and respect from other students)
I feel challenged	Rigorous expectations (e.g., teacher expectations and encouragement)
I feel prepared	Perseverance (e.g., focusing on a goal and overcoming problems when reaching a goal)
	Self-management (e.g., follow directions in class and not waiting until the last minute to do work)
	Self-efficacy (e.g., confidence in ability to understand concepts and complete assigned work)

Step 3: Developing business rules for creating the LCP index

Key consideration: Defining the overall LCP index using separate indices makes it possible to consider each separately and helps identify high priority areas for improvement.

1. Calculate average response within each component.
2. For each component (separately), determine whether the component exceeds 3.5, which represents a positive outcome.
3. Define a student to be LCP if they have positive outcomes for each component.
4. Average across students to determine the district-level LCP index.

Step 4: Providing evidence to inform key decisions

How do we weight aspects of preparedness?

Question: Should the three component scales in the Preparedness Index (perseverance, self-management, and self-efficacy) be weighted differently?

- Are the components equally important?

Evidence:

- Survey data: Contemporaneous correlations between scales and student outcomes showed that self-management and self-efficacy were better predictors than perseverance.
- Literature: Perseverance is the best predictor of long-term outcomes.

Decision: Weight each component equally.

How do we account for nonresponse bias?

Question: Should the district-level index include nonresponse weights?

- The usefulness of the LCP index depends on whether it represents all students.
- If only a select type of student responds, the resulting LCP index could be misleading.

Evidence: Examine sensitivity of results to nonresponse

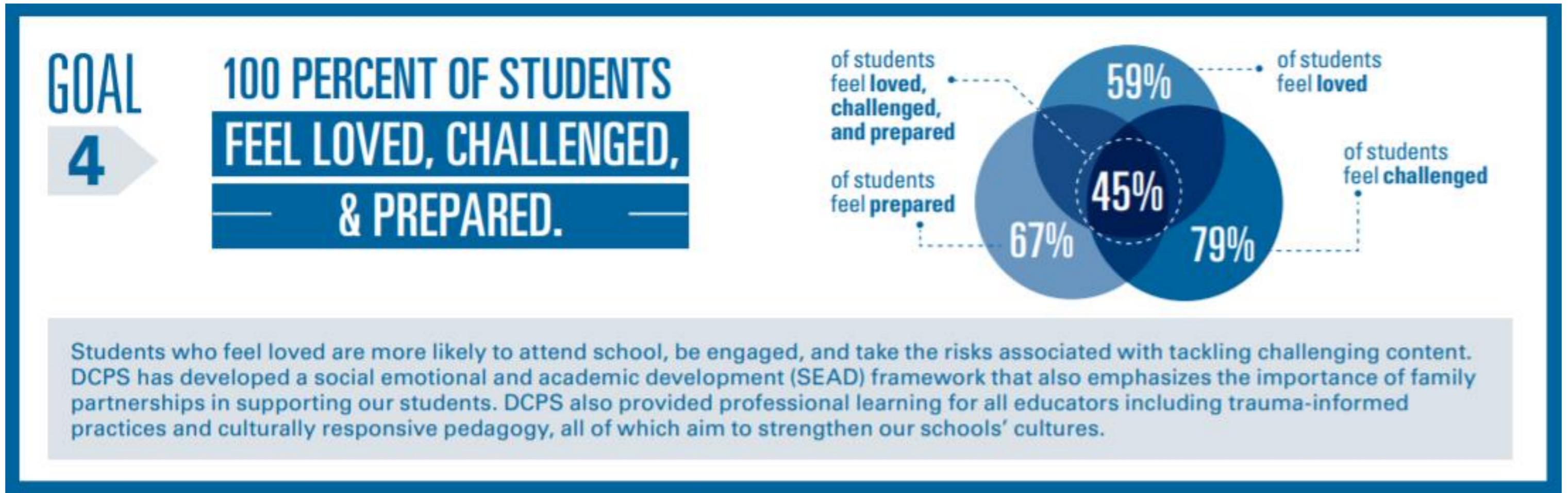
- Some evidence of differential response rates
- Overall index is not sensitive to using weights
- Limited low-effort responses

Decision: Weight all respondents equally

- Increased transparency to stakeholders

What we found

Results and DCPS's dissemination



Source: DCPS, 2018a.

Complementary past and future activities

- Previous coaching activities included the following:
 - Exploring properties of the teacher survey
 - Using natural language processing techniques to highlight themes in parent open-ended responses
- A future study will explore other uses of the SEL data to support DCPS in tracking progress toward its 2022 strategic goals.
 - It will study how SEL competencies predict key educational outcomes, such as grade transition and graduation.
 - It will investigate the extent to which reports of SEL skills change across grades.

Using Kindergarten Entry Assessments to Track City Progress in Promoting Reading Proficiency

A partnership with the School District of Philadelphia (SDP) and REL Mid-Atlantic

Jessica F. Harding, Mathematica

Mariesa Herrmann, Mathematica

Kristyn Stewart, School District of Philadelphia

Katherine Mosher, School District of Philadelphia

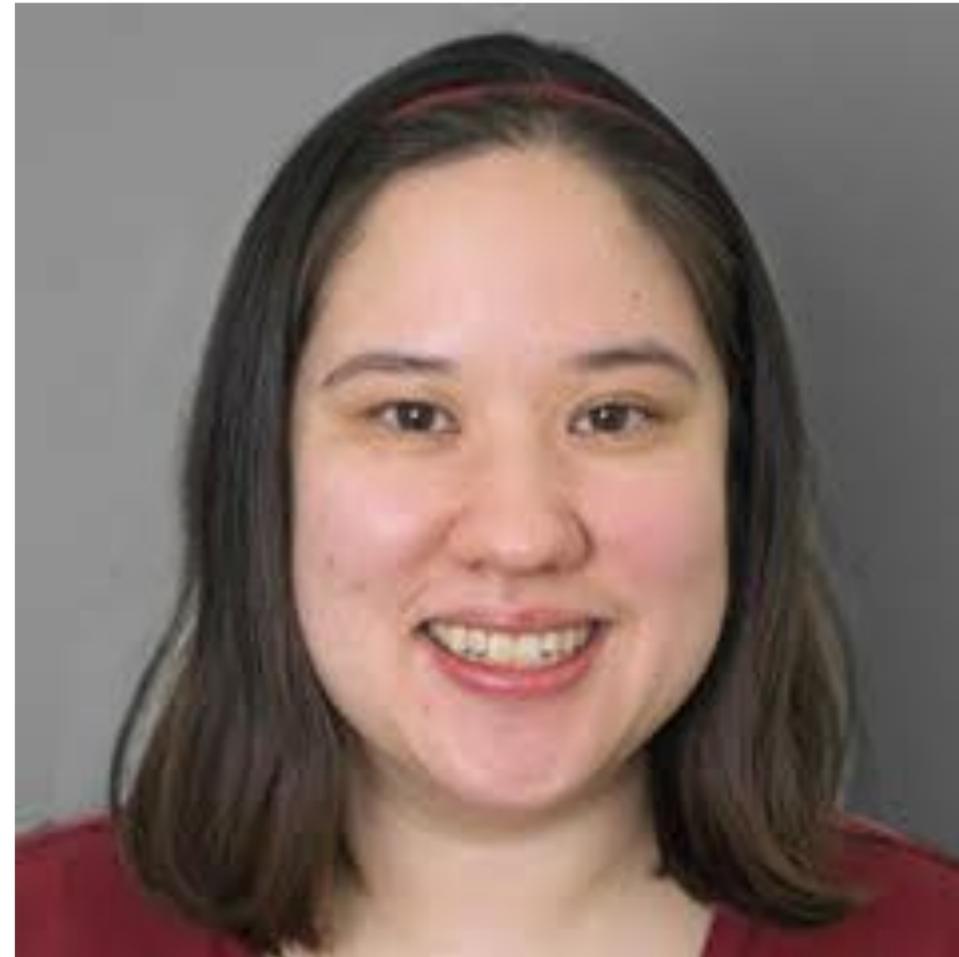
Elias Hanno, Burning Glass Technologies

Christine Ross, Mathematica

The REL Mid-Atlantic Presenters



Jessica Harding, Mathematica



Mariesa Herrmann, Mathematica

Purpose of the project



To help the School District of Philadelphia (SDP) track the progress of entering kindergartners in reading proficiently by the end of grade 3

- Kindergarten entry assessments have informed policymakers' decisions about early learning systems and identified children's skills to guide teaching (Regenstein et al., 2017).¹
- This project addresses a high-leverage need for using kindergarten entry assessments to track progress for later reading proficiency.
- Our partnership included the following:
 - Research to help select a threshold on Pennsylvania's Kindergarten Entry Inventory (KEI) for measuring whether kindergartners are on track for reading proficiency in grade 3
 - Coaching the district on selecting a threshold and providing code

¹ Regenstein, E., Connors, M., Romero-Jurado, R. I. O., & Weiner, J. (2017). *Uses and misuses of kindergarten readiness assessment results*. Chicago: The Ounce of Prevention Fund.

Research questions

Primary:

1. What is the relationship between scores on the KEI and the grade 3 Pennsylvania System of School Assessment (PSSA) in English language arts (ELA)?
2. What threshold score on the KEI most accurately predicts ELA proficiency on the grade 3 PSSA for the cohort?
 - What threshold score on the KEI has a prediction rate of at least 90 percent for the individual students who were not proficient on the PSSA?

Exploratory:

3. How do AIMSweb Reading scores in the spring of kindergarten to grade 3 relate to scores on the KEI and grade 3 PSSA in ELA?

Data for children who entered kindergarten in 2014–2015

1. KEI dimension scores (Emerging Academic Competencies and Learning Engagement Competencies)
2. PSSA reading scores and proficiency level
3. Enrollment records on school, gender, race and ethnicity, free and reduced price lunch, Individualized Education Program, English learner students
4. AIMSweb assessment scores and performance tiers

Cohort	2014-15	2015-16	2016-17	2017-18
2014-2015	Grade K	Grade 1	Grade 2	Grade 3
	Enrollment			
	KEI			
	AIMSweb			
				PSSA

Sample and methods

- Sample of kindergartners with both KEI and PSSA scores needed for analysis
 - 58 percent of kindergartners had KEI scores
 - 65 percent of kindergartners with KEI scores also had PSSA scores
- Weights were developed to make the sample with KEI and PSSA scores similar to the sample with KEI scores
- Data randomly split into two samples for analyses:
 - Sample used to select threshold
 - Sample used to test threshold

Results: Research question 1

EAC received greater weight in an index based on the two dimensions

What is the relationship between scores on the KEI and the grade 3 PSSA in ELA?

	Findings from logistic regression that includes both dimensions	
KEI dimensions	Coefficients	Average marginal effects
Emerging Academic Competencies (EAC)	1.147***	0.24***
Learning Engagement Competencies (LEC)	0.108	0.02

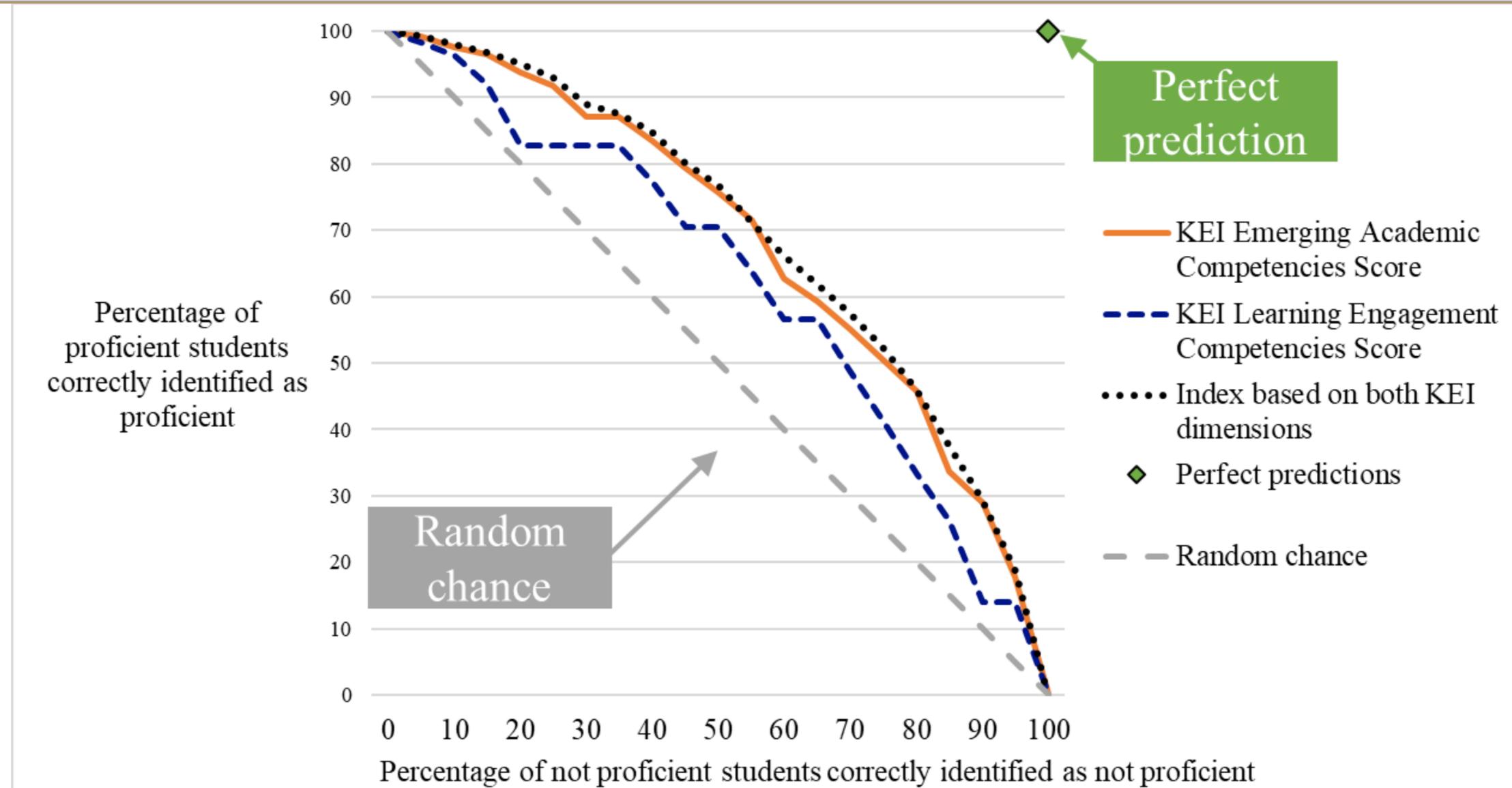
Statistical significance level: *** < .001

Number of students = 3,521

Results: Research question 2

EAC and index based on both EAC and LEC correctly predict proficiency for similar percentages of students

What threshold score on the KEI most accurately predicts reading proficiency on the grade 3 PSSA in ELA for the cohort?



Results: Research question 2

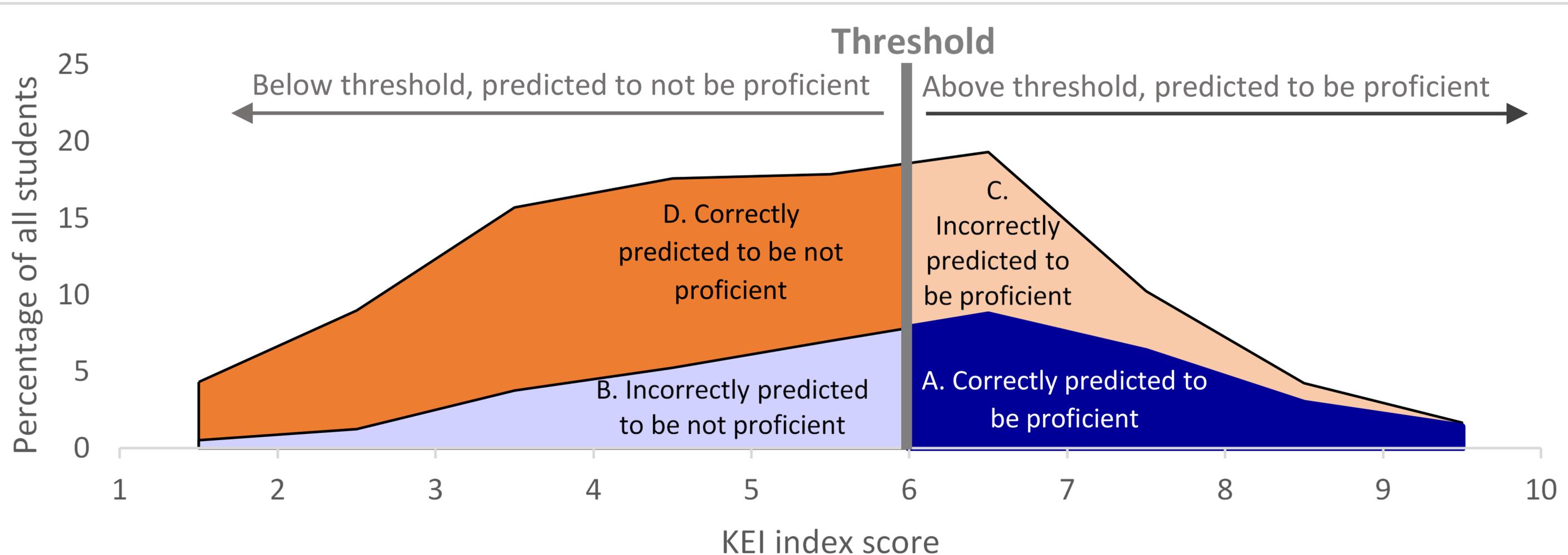
Threshold of 6 accurately predicted the actual grade 3 proficiency rate of students

Performance of threshold of 6 on KEI index of two dimensions

Data	Actual proficiency (percentage)	Predicted proficiency (percentage)	Difference between actual and predicted proficiency (percentage points)	Number of students
Used to set the threshold (in-sample)	37.2	36.6	0.6	3,521
Not used to set the threshold (out of sample)	36.9	36.7	0.2	1,508

Results: Research question 2

Threshold of 6 correctly predicted overall proficiency rate but made some errors in predictions for individual students



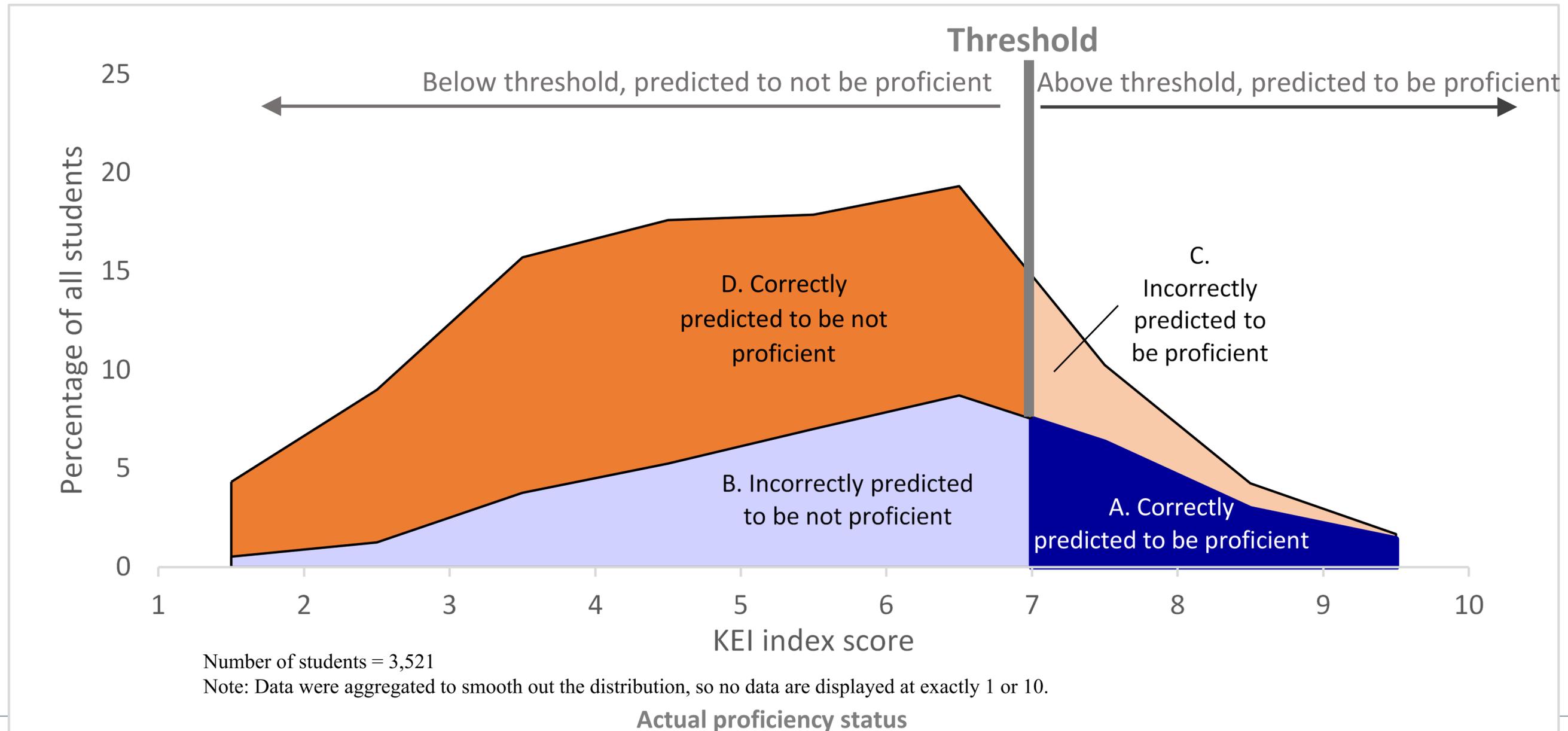
Notes: Data were aggregated to smooth out the distribution, so no data are displayed at exactly 1 or 10. Number of students = 3,521.

Not proficient

Proficient

Results: Research question 2

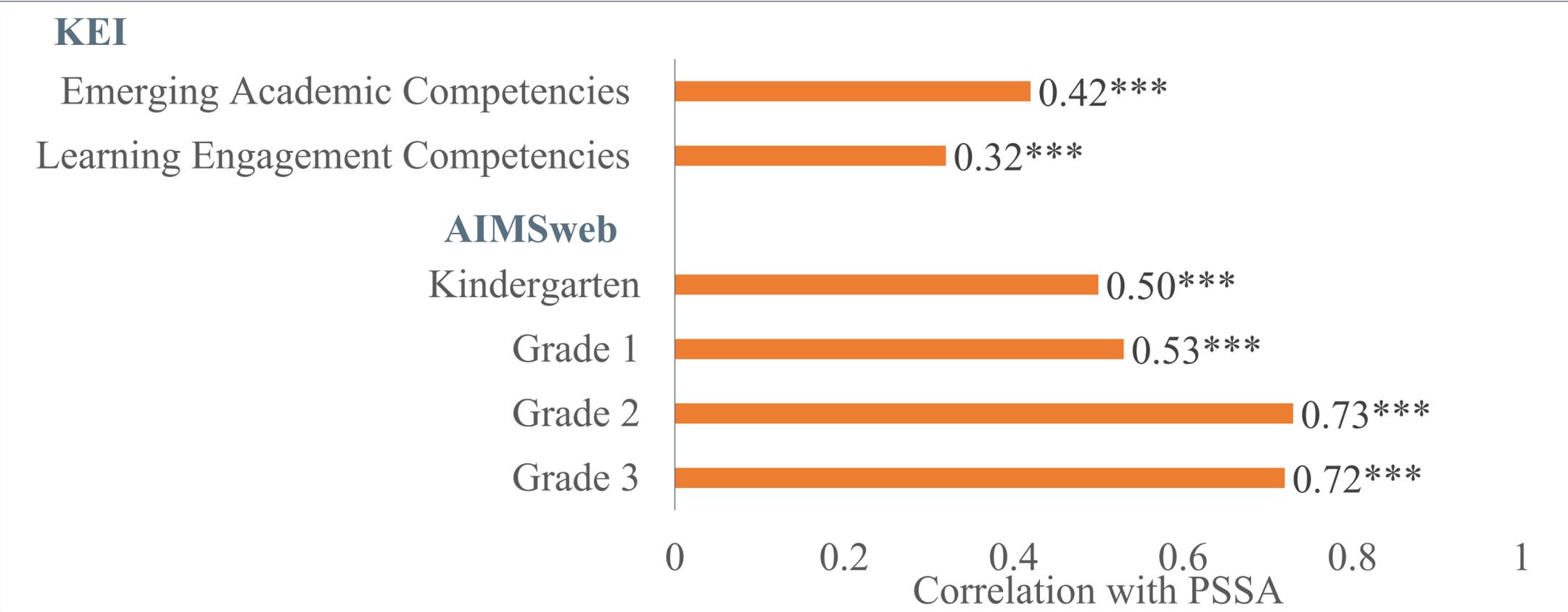
Threshold of 7 made correct predictions for 91 percent of students who were not proficient but only 29 percent of proficient students



Results: Research question 3

AIMSweb scores in later grades had higher correlations with PSSA scores in grade 3

How do AIMSweb Reading scores in the spring of kindergarten through grade 3 relate to children's scores on the KEI and grade 3 PSSA in ELA?



Notes: Correlations for spring AIMSweb scores were not weighted. Number of students with KEI correlations = 5,321. Number of students with AIMSweb correlations = 3,697 to 5,266, depending on whether students had scores on the AIMSweb in the grade and the PSSA. Statistical significance levels: *** < .001.

Limitations

- A sample with both KEI *and* PSSA scores might not represent the sample with KEI scores, even with weights.
- KEI threshold might not yield the same predictive power when applied to future cohorts if assessments or demographics of cohorts change.
- The study relies on the two validated dimensions of the KEI, but other KEI items or alternative assessments could have stronger relationships with PSSA proficiency.

Implications and next steps

- KEI, particularly Emerging Academic Competencies, is related to PSSA:
 - Possible to set a threshold that accurately predicts the cohort proficiency rate
 - Different threshold might be more appropriate for identifying individual students who need supports
- SDP plans to conduct additional analyses using data on more recent cohorts of kindergartners:
 - Reassess the accuracy of or update the KEI threshold
 - Assess the accuracy of thresholds on different measures
- States or districts might be able to use kindergarten entry assessments to do the following:
 - Track progress toward later reading proficiency across cohorts of kindergartners
 - Target supports to schools and students at risk of not reading proficiently

Measuring School Performance for Early Elementary Grades in Maryland

A partnership with the Maryland State Department of Education

Dallas Dotter, Mathematica

Lisa Dragoset, Mathematica

Cassandra Baxter, Mathematica

Elias Walsh, Mathematica

The REL Mid-Atlantic Presenter



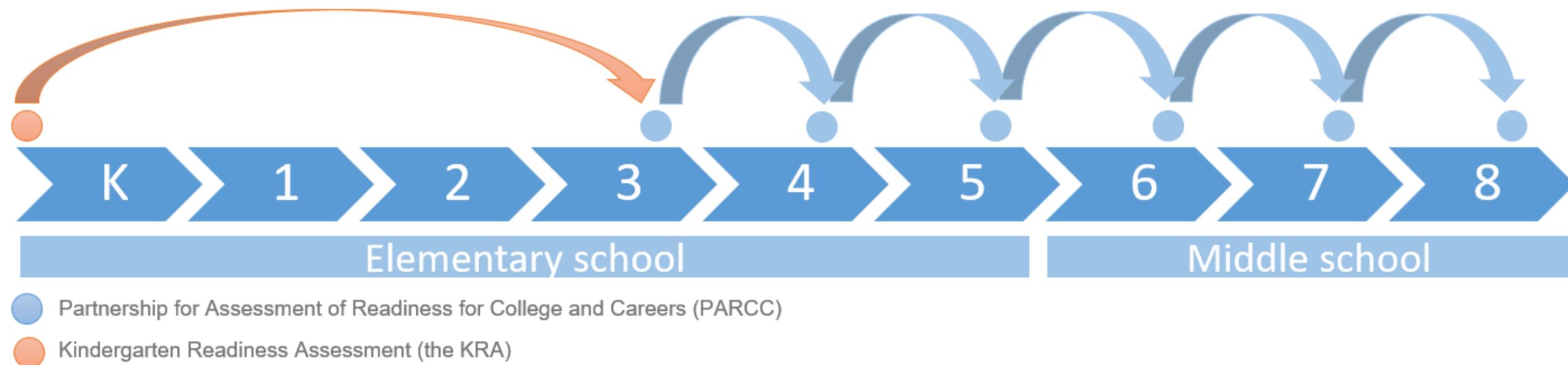
Dallas Dotter, Mathematica

Why this study?

Maryland wanted to add academic growth in early grades to its accountability framework.

The Kindergarten Readiness Assessment (KRA) provides an opportunity to measure growth from kindergarten entry to the end of grade 3.

Maryland State Department of Education (MSDE) partnered with the REL to study the feasibility of measuring schools' contributions to academic growth in reading and math from K to 3.



Study sample and student growth percentile (SGP) model



Students with valid 2014–2015 KRA scores and 2017–2018 grade 3 PARCC scores:

- 54,393 for math and 54,397 for reading (86 percent of all students with a KRA score)



914 schools were used to explore the statistical properties of the estimates.



SGPs were estimated separately for grade 3 reading and math using scaled scores and aggregated to the school level by calculating the mean SGP.

- Students who transferred schools between K to 3 partially contribute to estimates for each school attended
- Estimates account for KRA measurement error (ranked SIMEX method)

Key research questions

1. Does the growth model perform better using particular KRA subscore combinations?

KRA overall score	Language and literacy subscore	Mathematics subscore	Social foundations subscore	Physical well-being and motor development subscore
-------------------	--------------------------------	----------------------	-----------------------------	--

2. Are schools' K–3 growth estimates valid and precise relative to estimates used for accountability in later grades?
 - Validity: Is student academic performance, as measured by the two assessments, related?
 - Precision: Will estimates vary over time, holding assessments and schools' performance constant?
3. How does school size affect the precision of K–3 growth estimates?
4. How would administering the kindergarten assessment to a random subsample of students affect the precision of the growth estimate?

Overall score is a practical KRA score for baseline measure

- Estimated correlations between students' grade 3 PARCC scores and different combinations of their KRA subscores
- Weights estimated using cross-validation, regressing students' grade 3 scores on KRA subscores
- Overall KRA scores and weighted combinations of subscores perform about as well in predicting grade 3 achievement

KRA score	Correlations: 2014–2015 KRA and 2017–2018 grade 3 PARCC scores	
	Math	Reading
Overall scaled score	0.53	0.53
Same-subject domain score	0.53	0.48
Weighted average: math and reading scores	0.55	0.54
Weighted average: all domain scores	0.56	0.55

Source: Administrative data provided by the Maryland State Department of Education.

Modest relationship between students' KRA and PARCC scores

- Correlations between KRA and grade 3 scores are significantly lower than correlations between students' grades 3 and 4 scores
 - Also true comparing grades 3 and 6 scores, which have similar time between assessments
- Other studies comparing performance on kindergarten assessments and later achievement find similar correlations:
 - Duncan et al., 2007: Range between 0.10 to 0.68; ECLS-K shows 0.68 for math, 0.56 for reading
 - Hanno et al., 2019: 0.43 reading competency; 0.32 reading engagement

Grades and school years	Correlation coefficient between students' assessment scores	
	Math	Reading
K–3; 2014–2015 to 2017–2018	0.53	0.53
3–4; 2014–2015 to 2015–2016	0.86	0.82
3–4; 2016–2017 to 2017–2018	0.87	0.85
3–6; 2014–2015 to 2017–2018	0.82	0.77

Source: Administrative data provided by Maryland State Department of Education.

Schools' K–3 and grade 3–4 mean SGPs are equally precise

- Bootstrap procedure used to estimate within-school variance of mean SGP
 - Nested bootstrap: Sample students with replacement within schools
 - Re-estimate student SGPs
 - Construct new mean SGPs for schools
- 95 percent confidence intervals for schools' SGP constructed from bootstrapped distribution
 - Confidence interval widths virtually identical between school-level K–3 and grades 3–4 SGP

Grade and school year	Average confidence interval width (percentile points)	
	Math	Reading
K–3 SGP (2017/18)	12	13
Grades 3–4 SGP (2017/18)	13	13

Source: Administrative data provided by the Maryland State Department of Education.

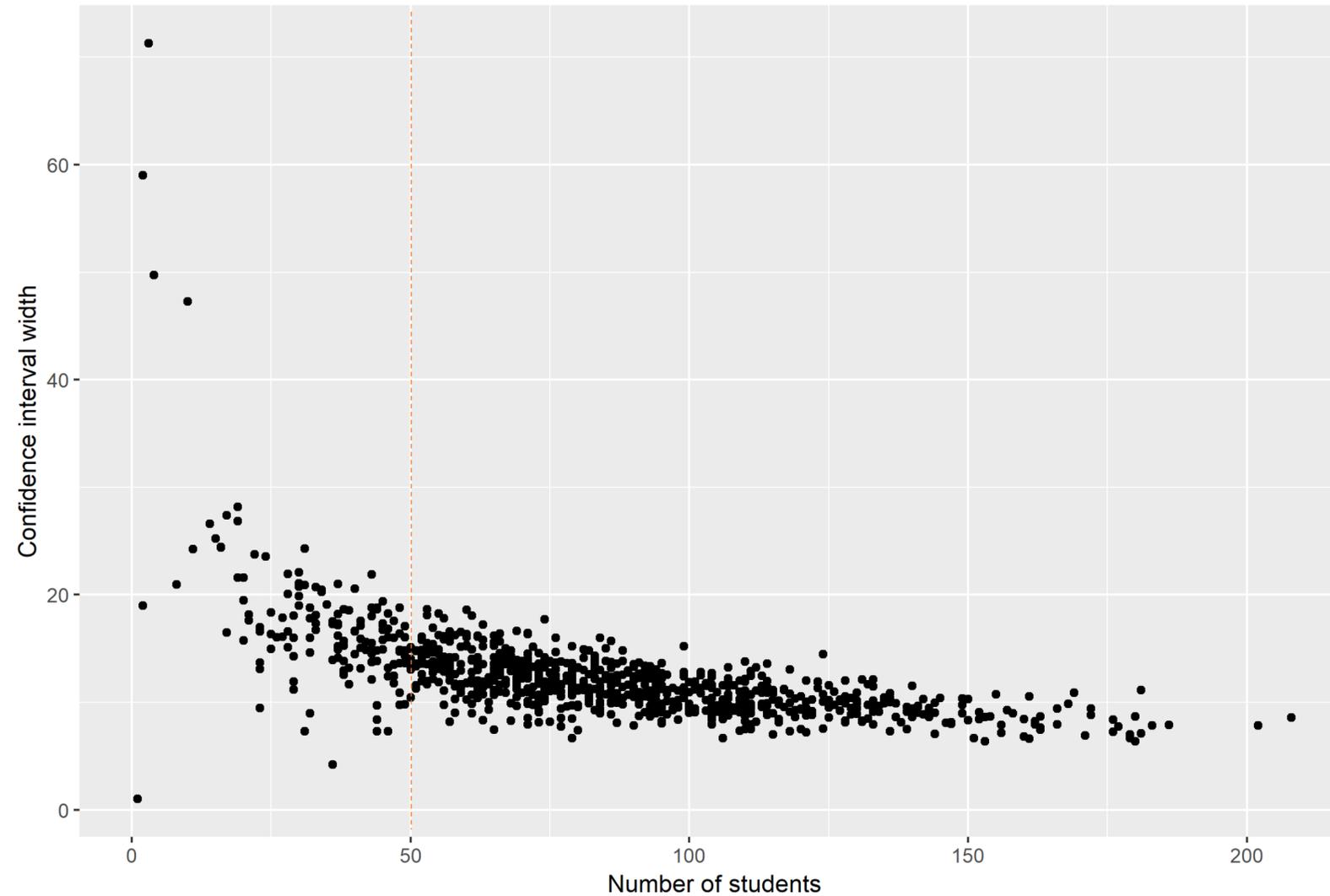
K–3 growth measure provides information complementary to growth in higher grades

Grades and school years	Correlation coefficient: Schools' K–3 and grade 3–4 growth estimates	
	Math	Reading
Grades 3–4 SGP (2015–2016)	0.45	0.52
Grades 3–4 SGP (2016–2017)	0.37	0.41
Grades 3–4 SGP (2017–2018)	0.40	0.37

Source: Administrative data provided by the Maryland State Department of Education.

- K–3 growth measure is moderately correlated with growth measure for 3–4
- Moderate correlation suggests that the information about K–3 growth is complementary to existing information, rather than redundant:
 - Schools showing strong growth in upper elementary grades do not necessarily show strong growth in early elementary grades

Estimates much less precise for some smaller schools



Note: Figure shows the relationship between school size (number of students contributing to a school’s K–3 growth estimate) and the confidence interval width for schools’ K–3 math growth estimates. Each dot represents a school.

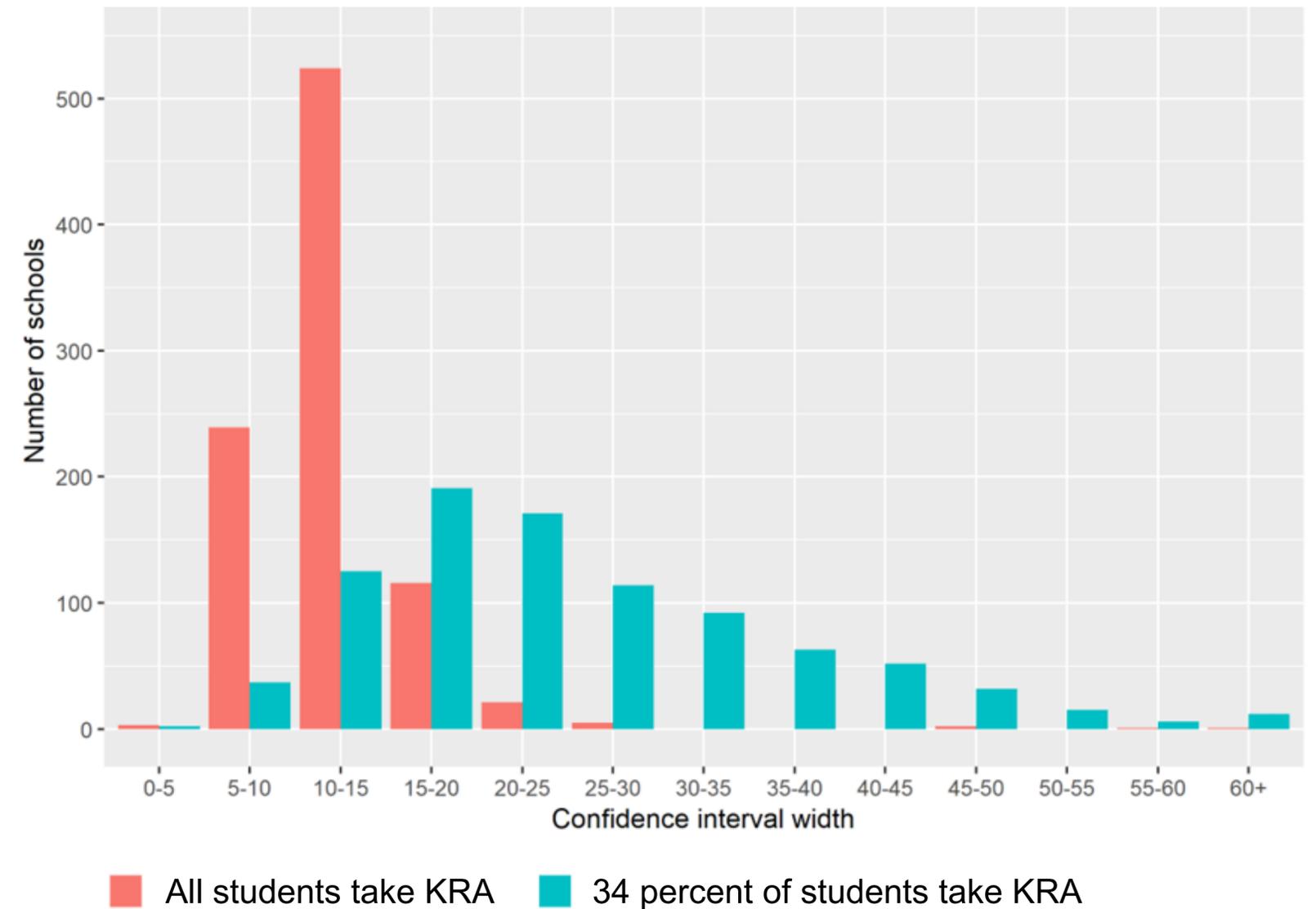
Partial KRA administration reduces precision

Average confidence interval widths, by KRA sampling proportion

Sampling percentage	Average confidence interval width (percentile points)	
	Math	Reading
Full sample (all students take KRA)	12	13
2016–2017 KRA sampling: 34 percent	25	26
2017–2018 KRA sampling: 35 percent	24	24
2018–2019 KRA sampling: 39 percent	23	23

Source: Administrative data provided by the Maryland State Department of Education.

Distribution of schools' confidence interval widths, by KRA administration



Limitations of the study

1. Findings and conclusions reflect only the students included in the analysis.
 - 14 percent of students with KRA scores did not have a valid grade 3 score
 - There are no KRA scores for grade 3 students entering Maryland public schools after kindergarten
2. Reliability for schools' K–3 growth estimates could not be studied.
 - This analysis could be conducted in future years (as data become available)
3. Unobserved student movement within school year could introduce bias.
 - Could be avoided if data were maintained on all schools attended by students each school year
4. Maryland will replace the PARCC with new statewide assessments.
 - A similar study can examine how the validity and precision of the estimates change

Implications and next steps

K–3 growth measure using KRA can provide new information on growth in early grades.

There is an opportunity for states with comprehensive kindergarten assessments to also understand how schools' early grade levels contribute to students learning.

The measure should be used with appropriate caution:

- Relatively low correlations suggest K–3 growth estimates less valid than estimates for grades 3–4
- Growth estimates are much less precise for smaller schools
- Precision is also a concern for schools in which the KRA is administered to only a portion of kindergartners

Findings are informing Maryland discussions about future use of K–3 growth measures.

Final Thoughts and Questions

Final thoughts

- Three examples of the value of working in RPPs to use existing data to develop new diagnostic measures
 - With DCPS: Can the District's SEL student survey be used to measure whether students feel loved, challenged, and prepared?
 - With SDP: Can the Kindergarten Entry Inventory be used as an early warning indicator of reading success in 3rd grade?
 - With MSDE: Is using the Kindergarten Readiness Assessment a valid way to measure school contribution to K-3 growth?
- Translating sophisticated research findings into changes in behavior
- Challenge of being both proactive and reactive in asking/answering research questions



Joy Lesnick
**Deputy Chief of Research,
Evaluation, and Academic
Partnerships**
**School District of
Philadelphia**

Questions?



Disclaimer

This work was funded by the U.S. Department of Education's Institute of Education Sciences (IES) under contract ED-IES-17-C-0006, with REL Mid-Atlantic, administered by Mathematica. The content of the presentation does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

<https://ies.ed.gov/ncee/edlabs/regions/midatlantic/>



Contact information

Brian Gill

BGill@mathematica-mpr.com

Tim Kautz

TKautz@mathematica-mpr.com

Kathleen Feeney

KFeeney@mathematica-mpr.com

Jessica Harding

JHarding@mathematica-mpr.com

Mariesa Herrmann

MHerrmann@mathematica-mpr.com

Dallas Dotter

DDotter@mathematica-mpr.com

References

- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246.
- Bland, J. M., & Altman, D. G. (1997). Cronbach’s alpha. *BMJ*, 314(7080), 572.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods Research*, 21(2), 230–258.
- District of Columbia Public Schools (DCPS). (2017). *DCPS strategic plan – A capital commitment 2017-2022*. https://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/DCPS%20Strategic%20Plan%20-%20A%20Capital%20Commitment%202017-2022-English_0.pdf
- District of Columbia Public Schools (DCPS). (2018a). *A capital commitment 2017-2022 – year 1 update*. <https://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/A%20Capital%20Commitment%202017-2022%20-%20Year%201%20Update.pdf>
- District of Columbia Public Schools (DCPS). (2018b). *DCPS 2018 Panorama survey results*. https://dcps.dc.gov/sites/default/files/dc/sites/dcps/page_content/attachments/2018PanoramaSurveyReport_0.pdf
- Dragoset, L, Baxter, C., Dotter, D., & Walsh, E. (2019). Measuring school performance for early elementary grades in Maryland. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic.
- Harding, J. F., Herrmann, M. A., Hanno, E. S., & Ross, C. (2019). *Using kindergarten entry assessments to measure whether Philadelphia’s students are on-track for reading proficiently*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic.
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4), 451–464.
- Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success* (OECD Education Working Papers No. 110). OECD Publishing.
- Stevens, J.P. (2012). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge.

Appendix 1 - Using Student-Level Social and Emotional Learning Data to Develop a “Loved, Challenged, and Prepared” Measure

Overview of the student topics used in building the index

Topic	# of items	
	Grades 3 to 5	Grades 6 to 12
Sense of belonging* - connections to adults at school and respect from other students	4	5
Rigorous expectations* - teacher expectations and encouragement	5	5
Perseverance - focusing on a goal and overcoming problems when reaching a goal	4	5
Self-management - follow directions in class and not waiting until the last minute to do work	5	5
Self-efficacy - confidence in ability to understand concepts and complete assigned work	5	5

Steps for validating survey

1. Prespecify validity criteria.
2. Calculate Cronbach's alpha (a measure of reliability) and conduct analyses to estimate the number of underlying factors per scale.
 - Ensure that each scale reliably captures a single factor.
3. Conduct an initial confirmatory factor analysis.
 - Assess overall fit.
 - Investigate whether scales capture distinct factors.
4. Revise scales as needed based on results.
5. Confirm that final scales meet prespecified criteria.
 - Distinct factors (discriminant validity)

How do we adjust the index for low-effort respondents?

Question: How do we identify low-effort respondents and their resulting impact or bias on the results?

- Straightlining
- Rapid completion

Evidence: Examine straightlining patterns.

- The overall index is not sensitive to excluding them (difference is 0.4 out of 100).

Decision: Include students who straightline in the index.

- Some students might be providing sincere responses.

Appendix 2 - Research Partnerships on Using Kindergarten Entry Assessments to Track City Progress in Promoting Reading Proficiency

Students' characteristics differ slightly for those with and without KEI/PSSA data

Student characteristics	K enrollment sample	KEI sample	KEI and PSSA sample
Number of students	9,317	5,398	3,521
Female	48%	48%	49%
Non-Hispanic Black	49%	48%	46% ^a
Hispanic	21%	20%	22% ^b
Non-Hispanic White	14%	16% ^a	15%
Other race or ethnicity	16%	17%	17%
Free or reduced-price lunch	66%	68%	68%
English learner	10%	10%	11%
Special education	8%	8%	5% ^{a, b}
EAC	-	2.40	2.44 ^b
LEC	-	2.52	2.55 ^b

^a Substantially different (> 0.05 SD) relative to K enrollment sample.

^b Substantially different (> 0.05 SD) relative to KEI sample.

Because SDP is interested in predicting proficiency for children enrolled in kindergarten, we weight the analysis samples to reflect the characteristics of the K enrollment sample.

Weights make samples more similar

Student characteristics	Unweighted K enrollment sample	Weighted KEI sample	Weighted KEI and PSSA sample
Number of students	9,317	5,398	3,521
Female	48%	48%	48%
Non-Hispanic Black	49%	48%	48%
Hispanic	21%	20%	20%
Non-Hispanic White	14%	16%	16%
Other race or ethnicity	16%	17%	16%
Free or reduced-price lunch	66%	68%	68%
English learner	10%	10%	10%
Special education	8%	8%	8%
EAC	-	2.40	2.40
LEC	-	2.52	2.52

None of the differences relative to the kindergarten enrollment sample or relative to the KEI sample were substantially different (> 0.05 SD).

Results: Research question 3, robustness check

AIMSweb findings were similar among sample of students who had scores on all of the assessments

Spring AIMSweb

Kindergarten

0.49***

Grade 1

0.53***

Grade 2

0.70***

Grade 3

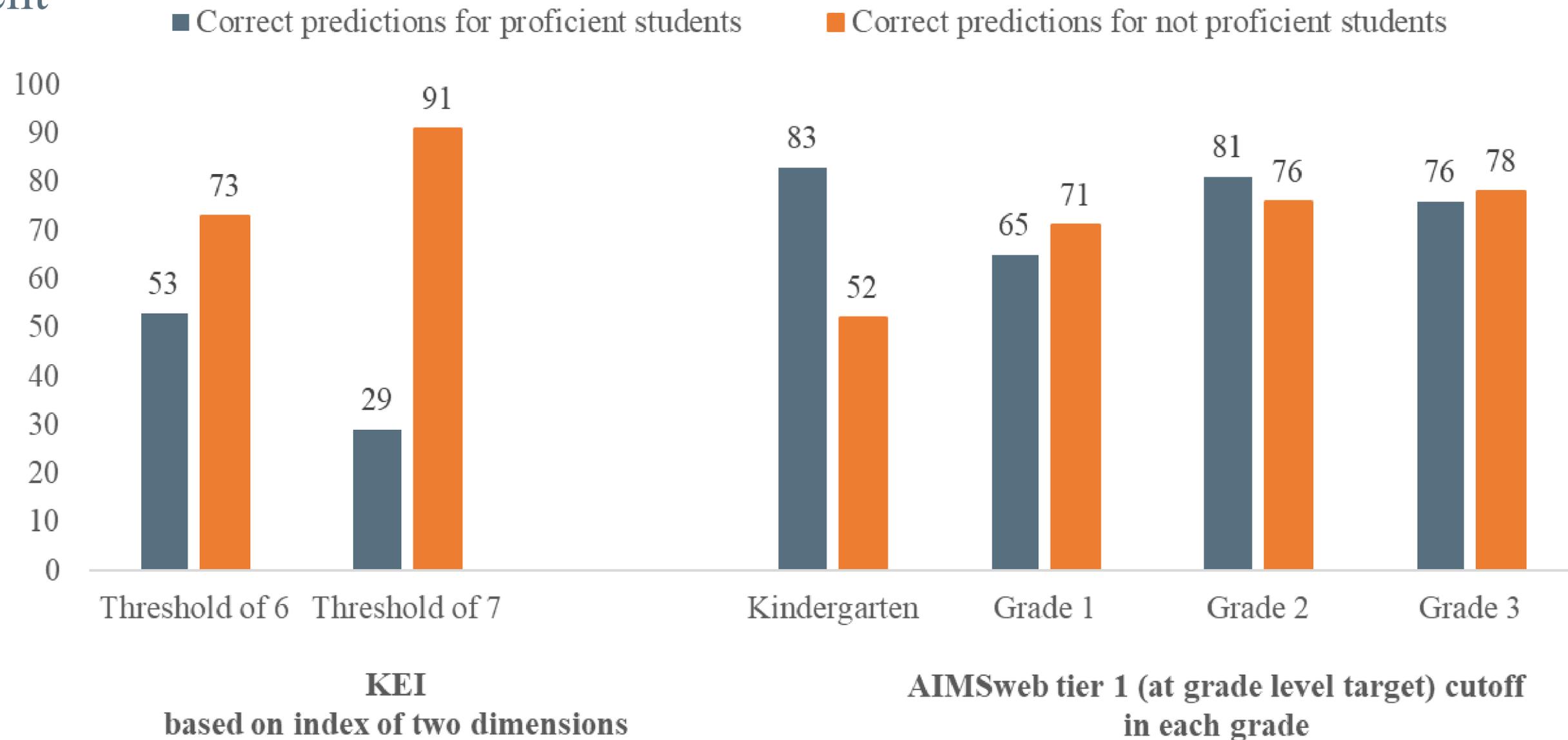
0.72***

0 0.2 0.4 0.6 0.8 1
Correlation with PSSA

Notes: Correlations for AIMSweb scores were not weighted. Number of students = 2,359. Statistical significance levels: *** < .001.

Results: Research question 3

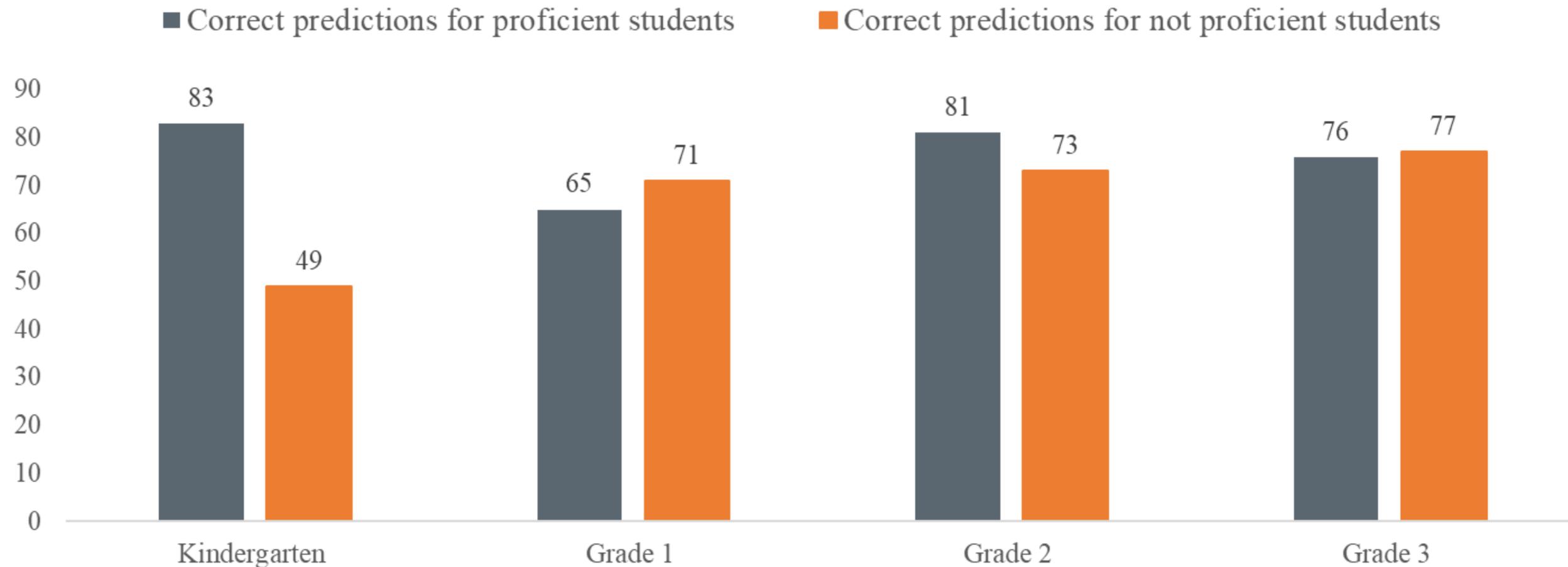
AIMSweb tier 1 cutoff correctly predicted more than 52 percent of the students who were not proficient



Notes: Analyses for spring AIMSweb scores were not weighted. Number of students with KEI correlations = 5,321. Number of students with AIMSweb correlations = 3,697 to 5,266 depending on whether students had scores on the AIMSweb in the grade and the PSSA.

Results: Research question 3, robustness check

AIMSweb findings were similar among sample of students who had scores on all assessments



AIMSweb Tier 1 (at grade level target) cutoff in each grade

Notes: Analyses for spring AIMSweb scores were not weighted. N = 2,358.

Results: Research question 3, supplementary findings

AIMSweb tier 1 cutoff was less accurate than KEI threshold for predicting overall proficiency rate

Assessment and threshold	Actual proficiency (percentage)	Predicted proficiency (percent)	Absolute difference between actual and predicted proficiency (percentage points)	Number of students
KEI index threshold (fall of kindergarten)				
6.00	37	37	1	3,521
7.00	37	17	21	3,521
AIMSweb tier 1 cutoff (spring of each grade)				
Kindergarten	37	61	23	4,696
Grade 1 AIMSweb tier 1	36	42	6	4,174
Grade 2 AIMSweb tier 1	37	45	8	5,027
Grade 3 AIMSweb tier 1	37	42	5	5,266

Notes: The difference between actual and predicted proficiency might not equal the difference between the actual and predicted proficiency because of rounding. Correlations for AIMSweb scores were not weighted.