

Validity Evidence for the STAR Assessments and Smarter Balanced Summative Assessments

Washington State School Directors Association (WSSDA)
2015 Annual Conference * Bellevue, WA

Jack B. Monpas-Huber, Ph.D.

Director of Assessment, Student Information & Highly Capable Program

jack_monpas-huber@msd25.org



Acknowledgments

Dr. Becky Berg
Cinco Delgado

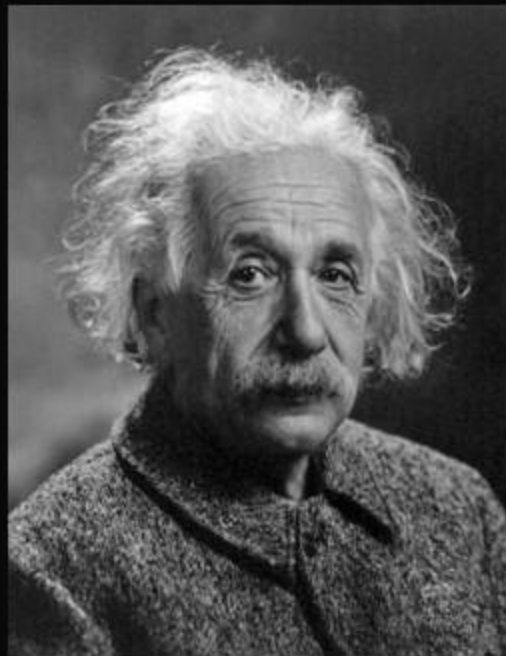
Teachers on Special Assignment

Marysville Administrative Team



Pierre Guerts
Eric Stickney
Jon Doney

The Renaissance Learning logo is contained within a dark blue rectangular box. It features the word 'RENAISSANCE' in white, uppercase, sans-serif font, followed by a white stylized swoosh that curves upwards and to the right. To the right of the swoosh is the word 'LEARNING' in a smaller, white, uppercase, sans-serif font, followed by a small trademark symbol (™).



I have no special talent. I am only passionately
curious.

(Albert Einstein)

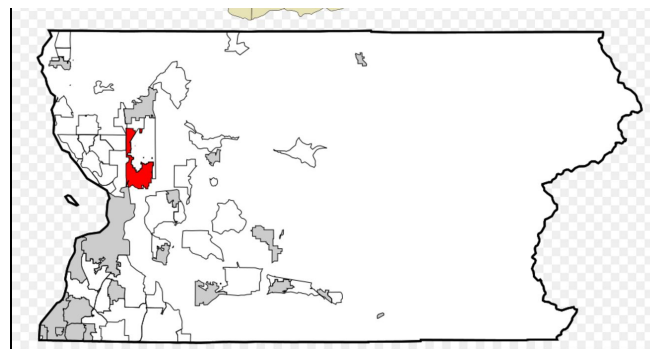
izquotes.com

Some context. . .

Marysville School District

November 2015 enrollment: 10,877

% free and reduced lunch: 46.2



- Need to increase student achievement and graduation and close gaps
- Need common district assessments, data
- Implemented STAR Assessments in Fall 2014
- First common district assessment, K-12
- Joined DIBELS and Fountas & Pinnell as elementary reading assessment

Reactions to the STAR and Smarter Balanced testing...

“A computer is not a valid way to assess reading. They’re just pressing buttons.”

“I saw a kid click ‘C’ all the way through..”

“There were so many tech glitches that it’s not a valid test. I don’t have any faith in the data.”

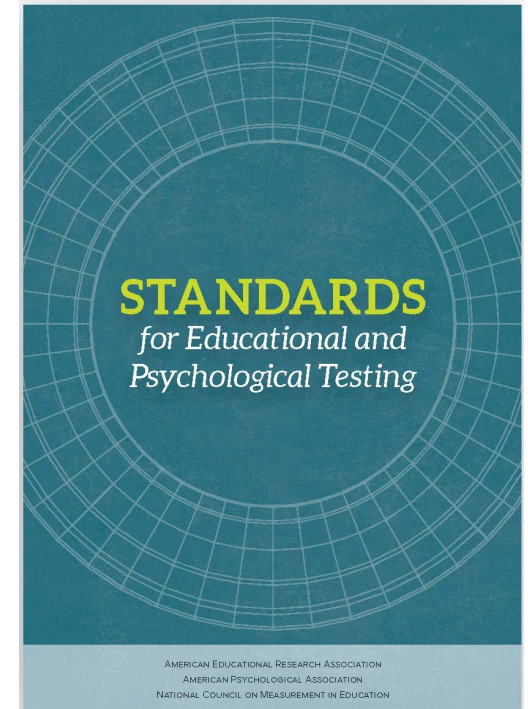


Turn and talk: How would you respond to these challenges?

I call these “validity threats”

Validity...

- “is the most **fundamental** consideration in developing and evaluating tests”¹
- is based on intended **purpose** of assessments^{1,2}
- is accumulating **evidence** of validity to support intended uses^{1,2,3}
- is **scientific** investigation of test score meaning^{1,3}, is “validity studies”
- is more than the **content** of the test items, is analyzing the **data** from the items and overall scores^{1,2,3}



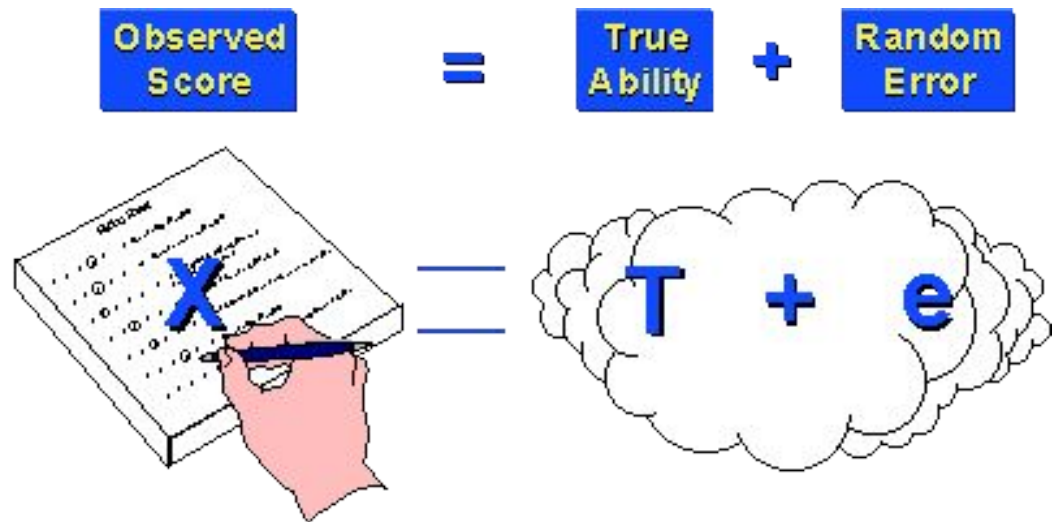
¹American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

² Kane, M.T. (2006). Validation. In: Brennan, F. L. (Ed.), *Educational measurement* (4th ed., pp.17-64.). Westport, CT: American Council on Education.

³ Messick, S. (1989). Validity. In Linn, R.L. (Ed.), *Educational measurement* (3rd ed., pp. 13-103.). New York: American Council on Education.

Validity evidence: What we know about reliability

At the individual level:



At the group level:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

Validity evidence: What we know about reliability

If most kids were just guessing....

Still get a distribution of scores, but it would be **random noise**

Score variance would be mostly **error variance**

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

If most kids were trying....

Would get a distribution of scores, and it would be mostly **true achievement**

Score variance would be mostly **true score variance**

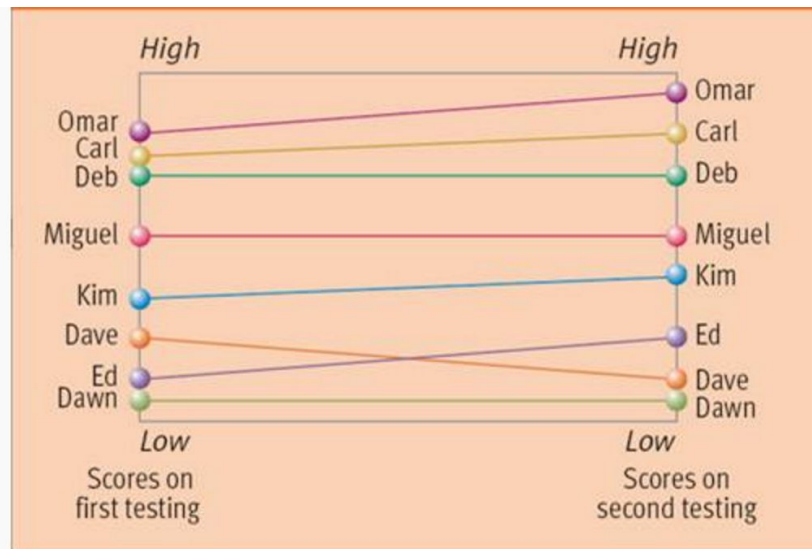
$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

Validity evidence: Correlations as reliability

One way to assess reliability is through the test-retest correlation

Correlation is a measure of how consistently two assessments sort out the same examinees

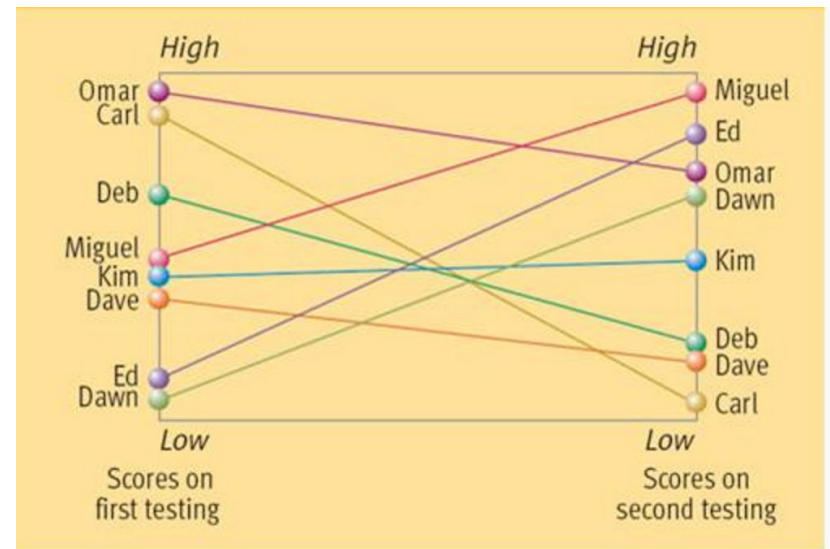
HIGH RELIABILITY



© 2007 Thomson Higher Education

STRONG CORRELATION

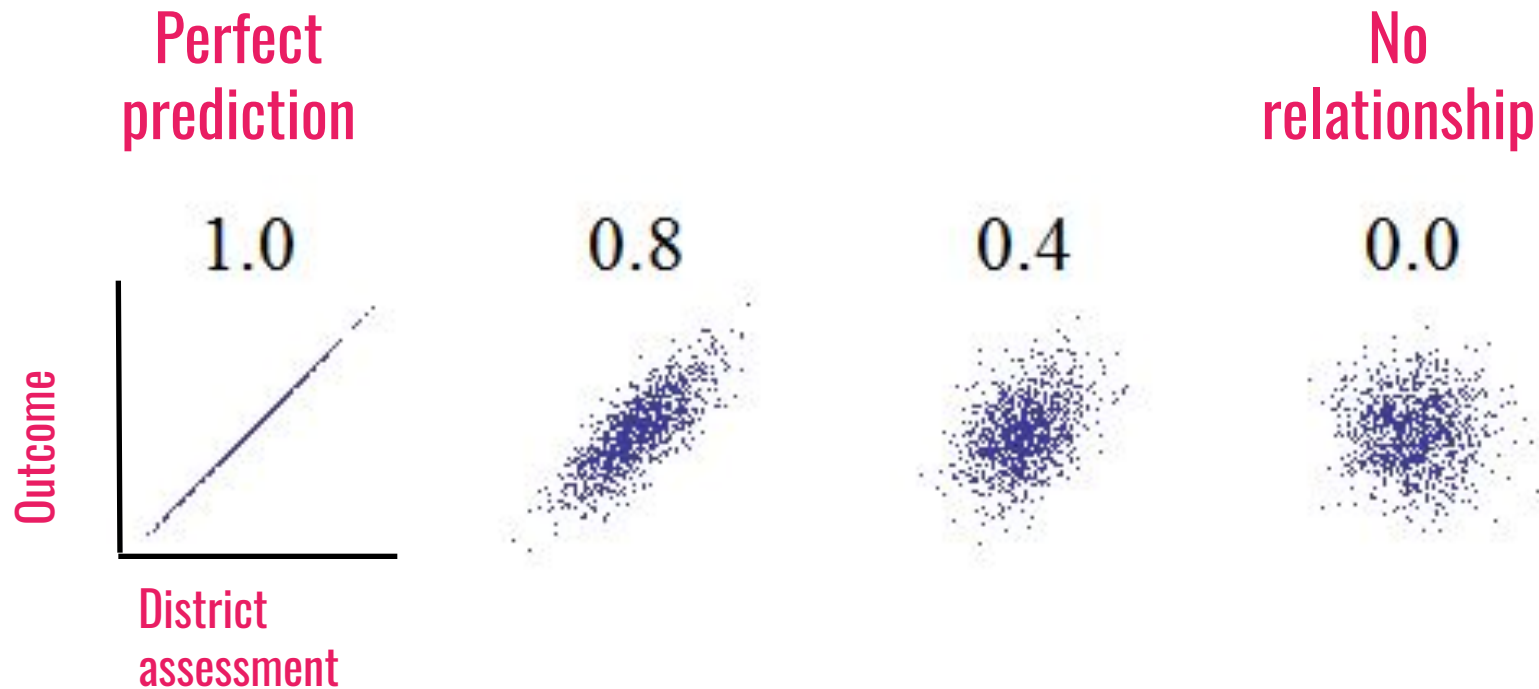
LOW RELIABILITY



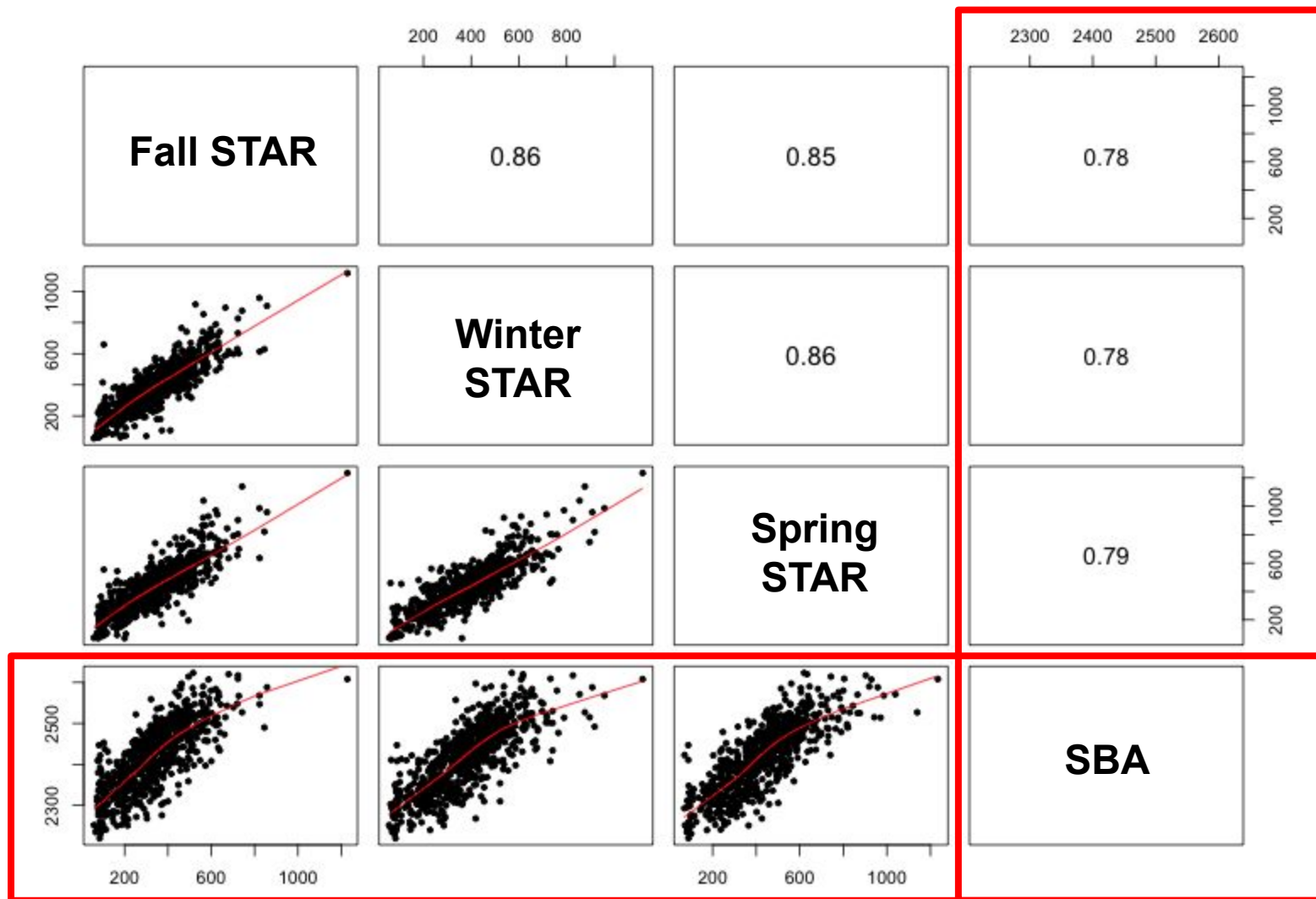
WEAK CORRELATION

Validity evidence: Correlations between STAR and Smarter Balanced assessments

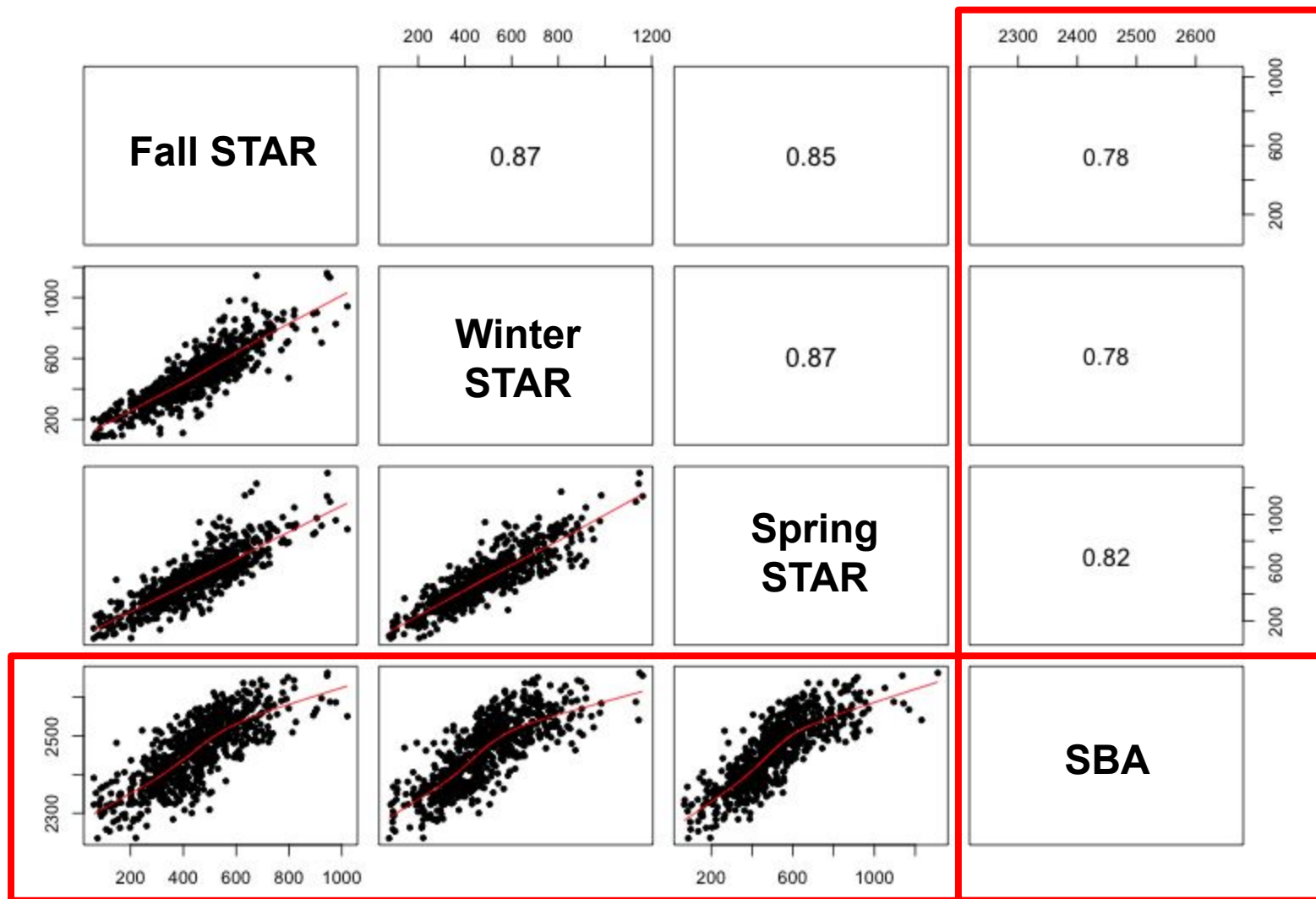
So what's a strong correlation look like?



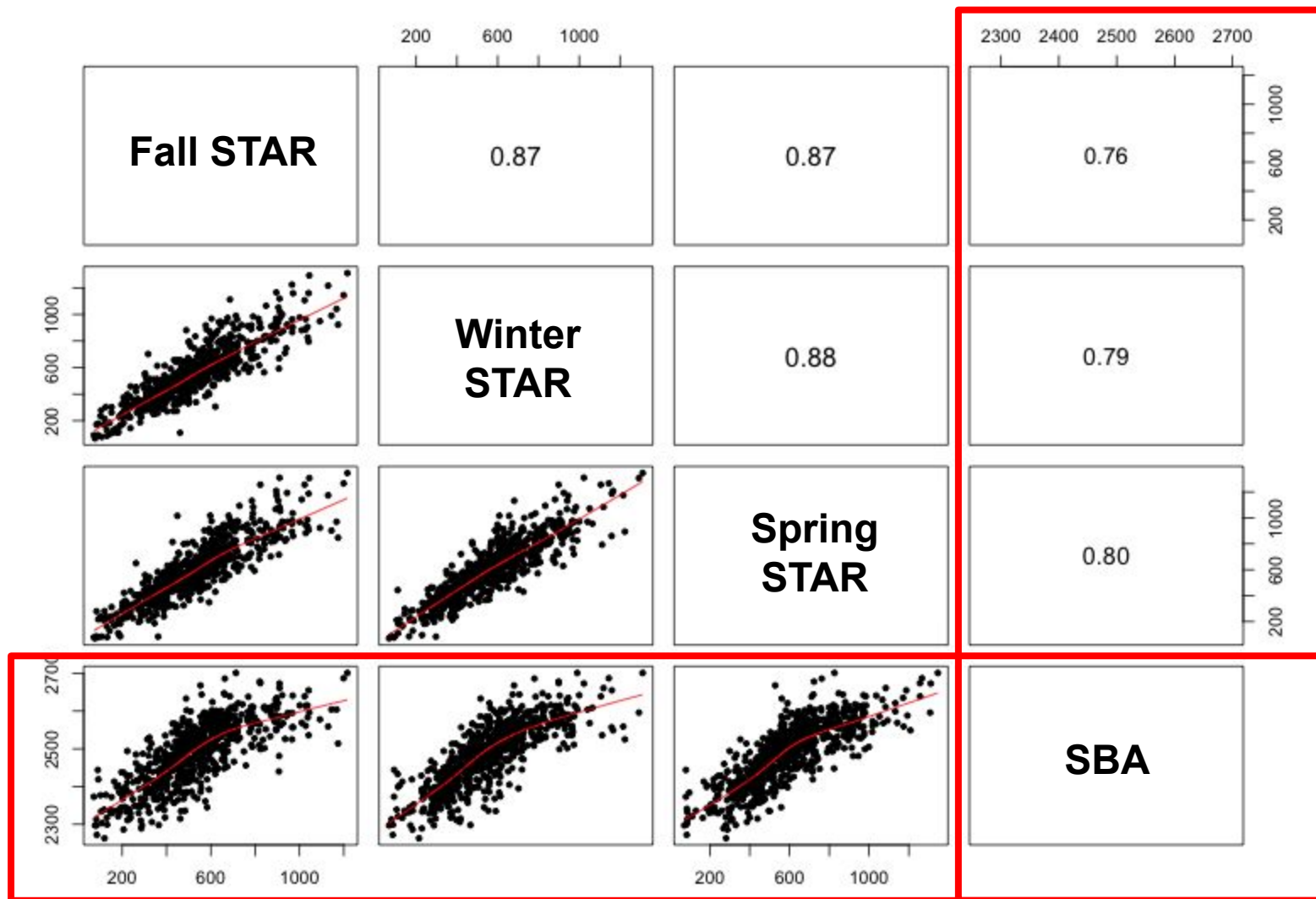
STAR Reading-SBA ELA Correlations - Grade 3 (N=709)



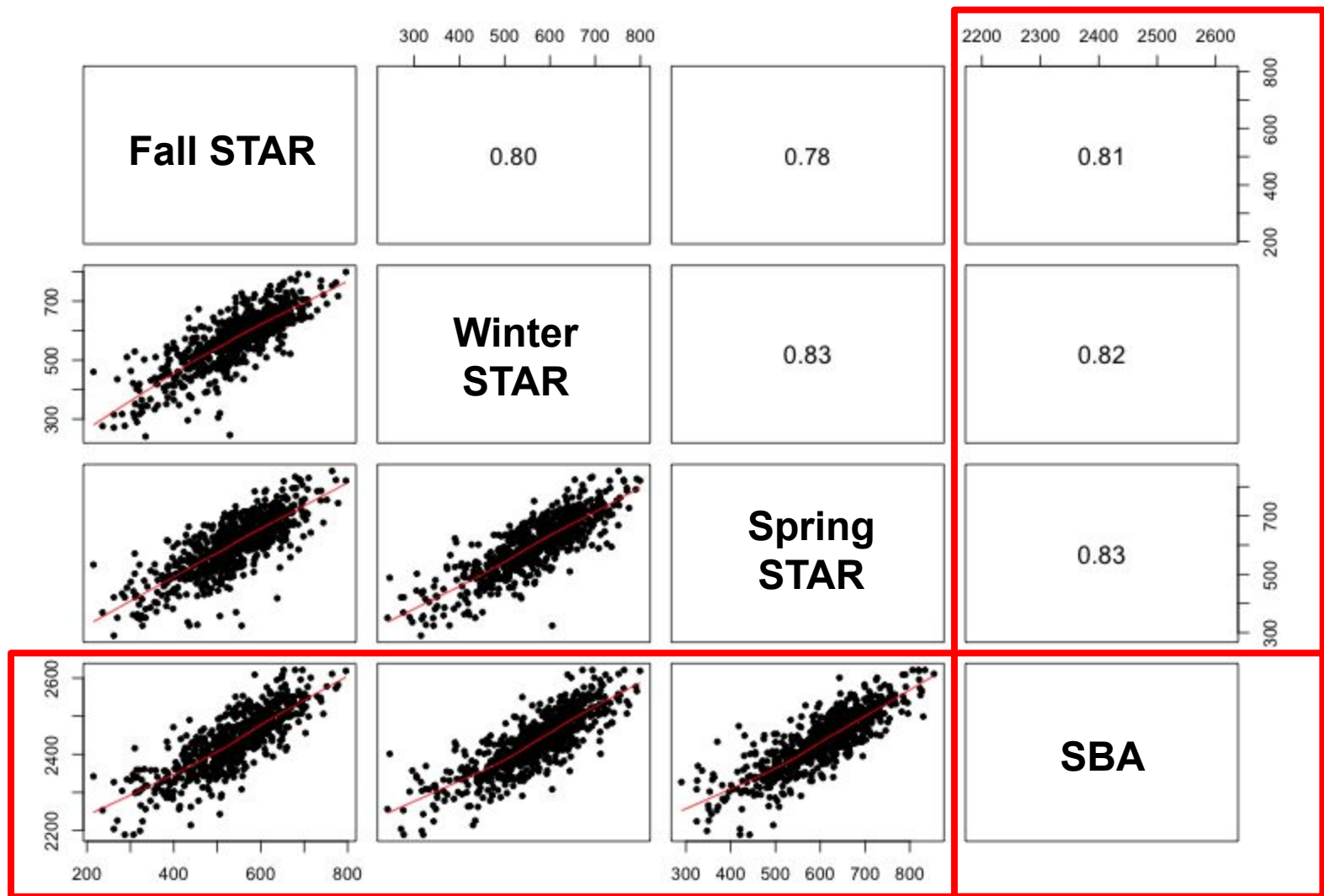
STAR Reading-SBA ELA Correlations - Grade 4 (N=690)



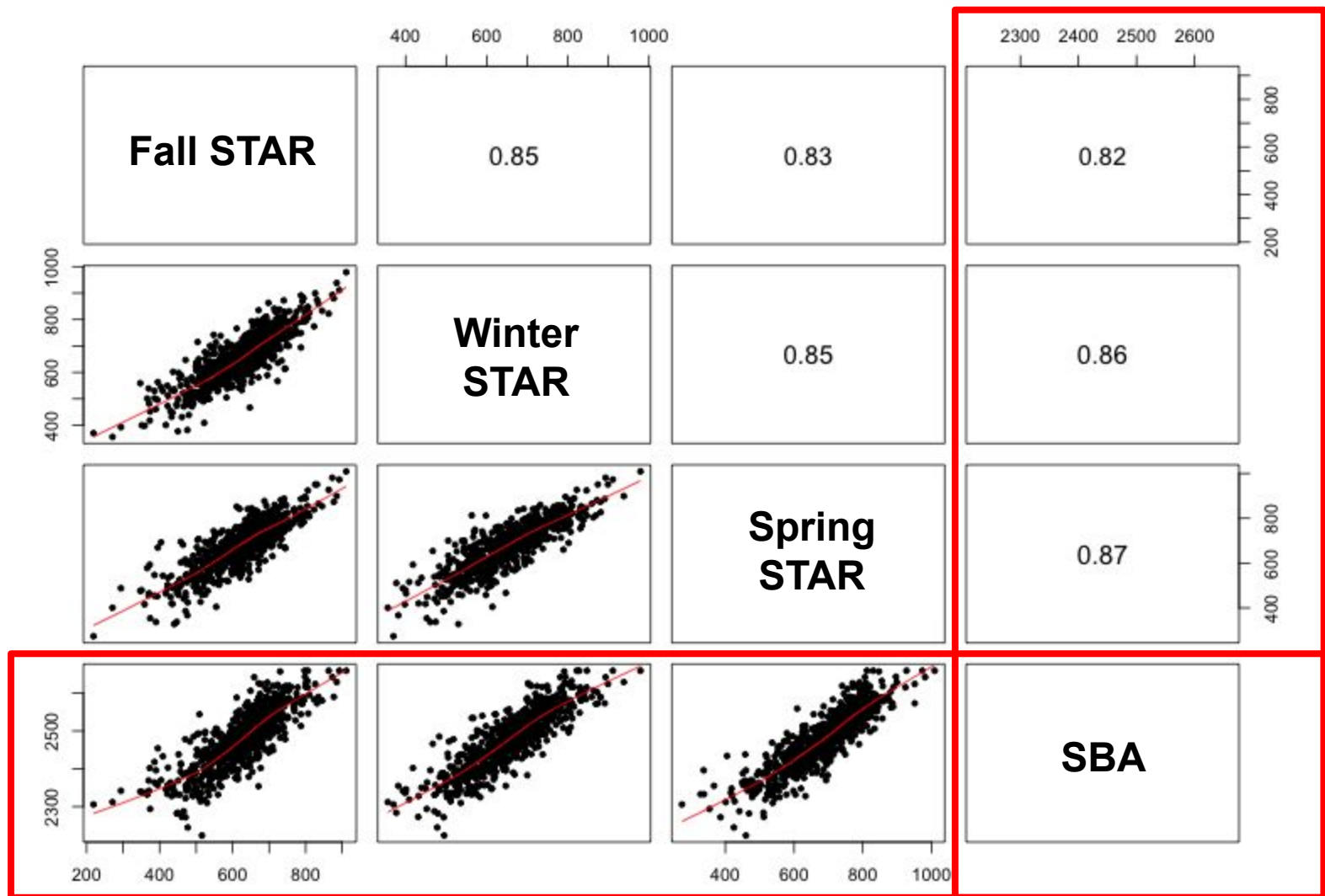
STAR Reading-SBA ELA Correlations - Grade 5 (N=697)



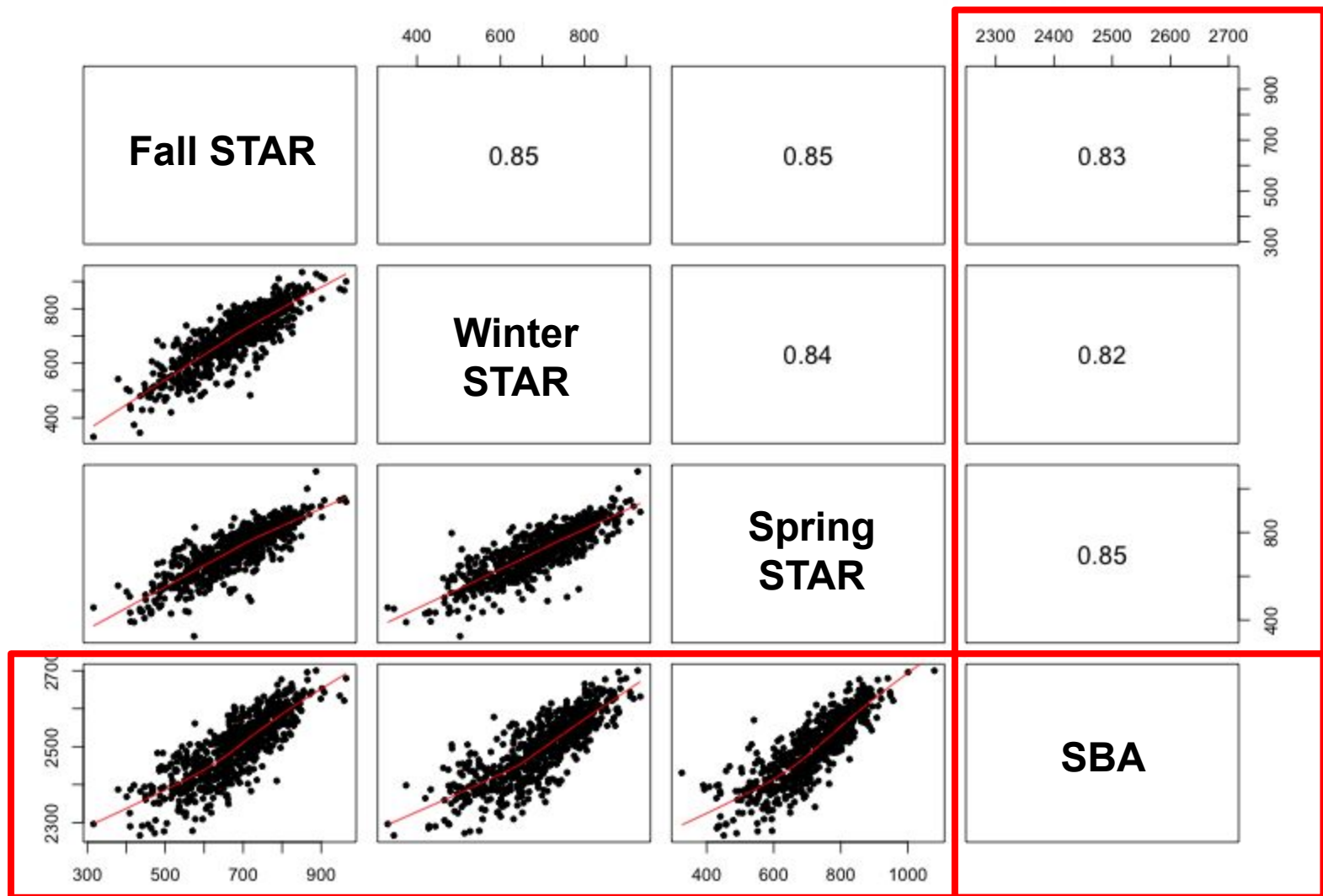
STAR Math-SBA Math Correlations - Grade 3 (N=703)



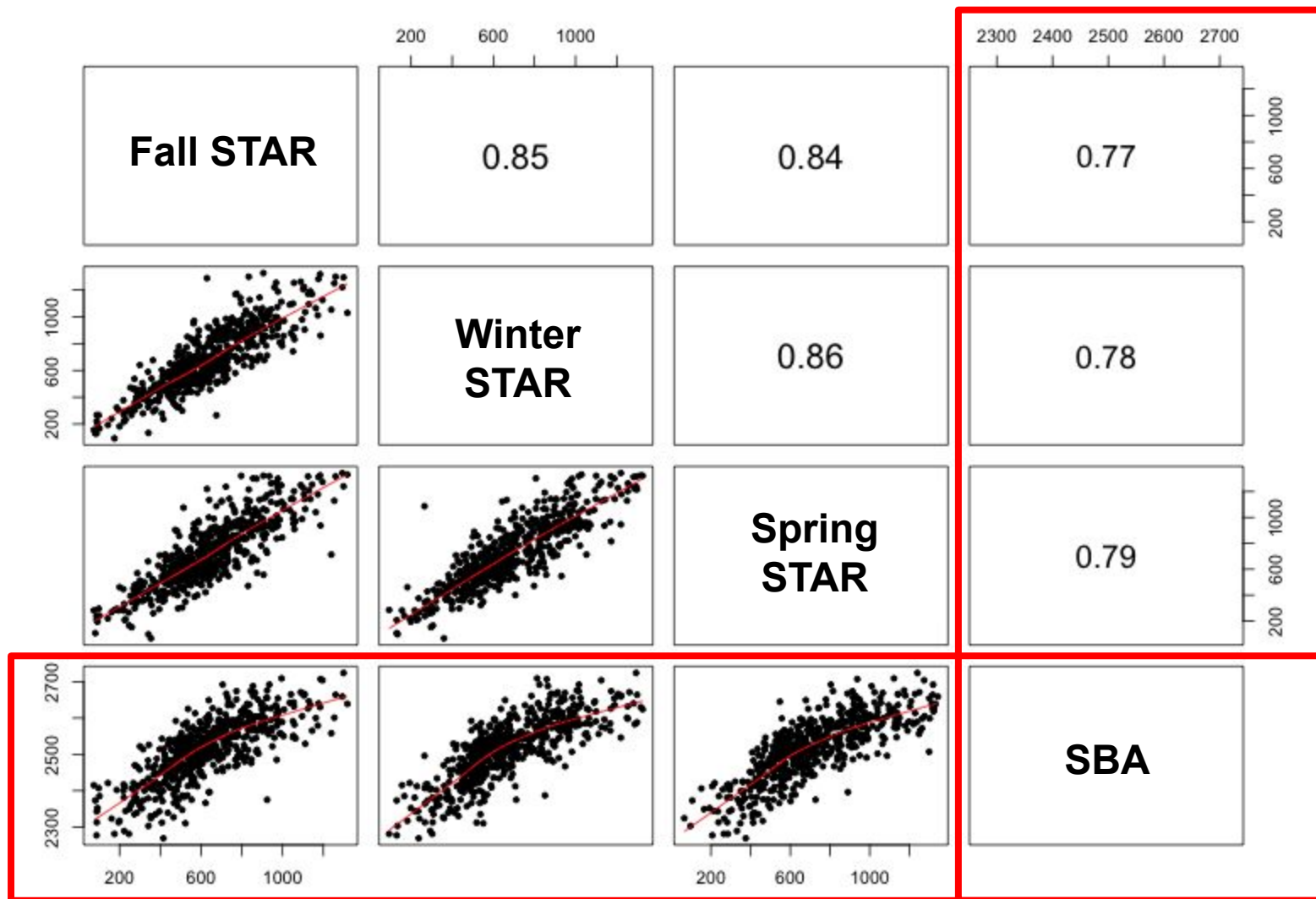
STAR Math-SBA Math Correlations - Grade 4 (N=706)



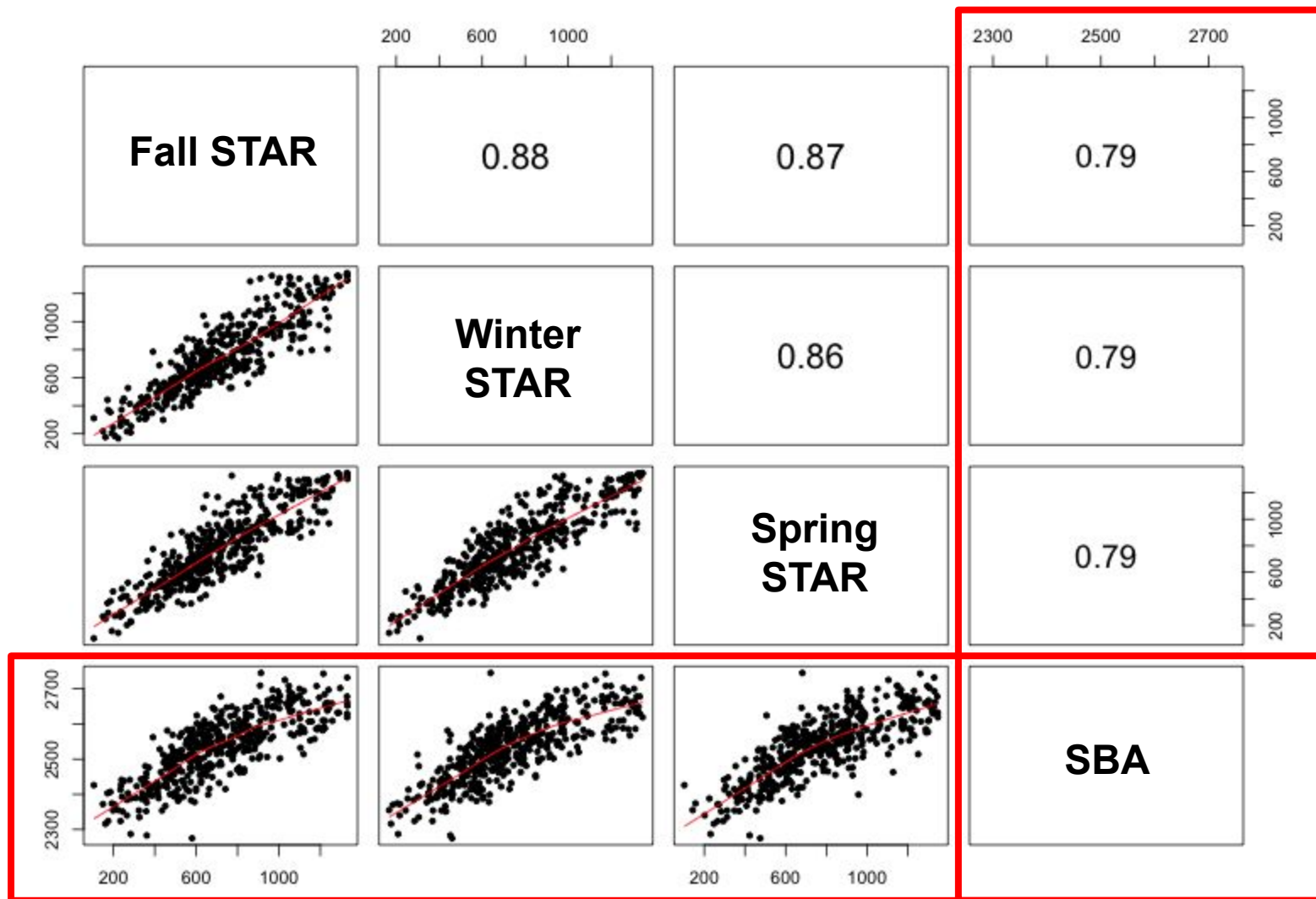
STAR Math-SBA Math Correlations - Grade 5 (N=660)



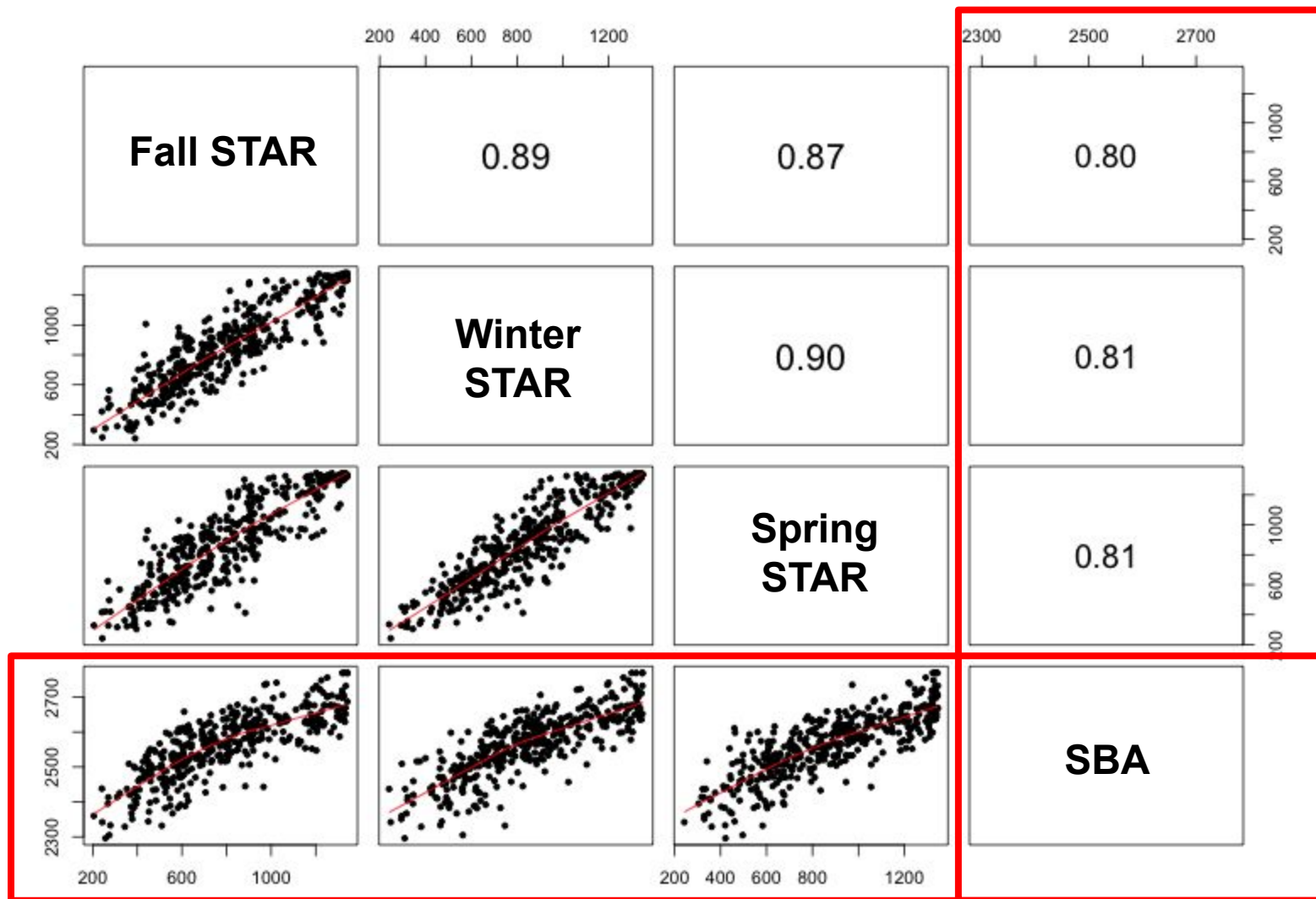
STAR Reading-SBA ELA Correlations - Grade 6 (N=567)



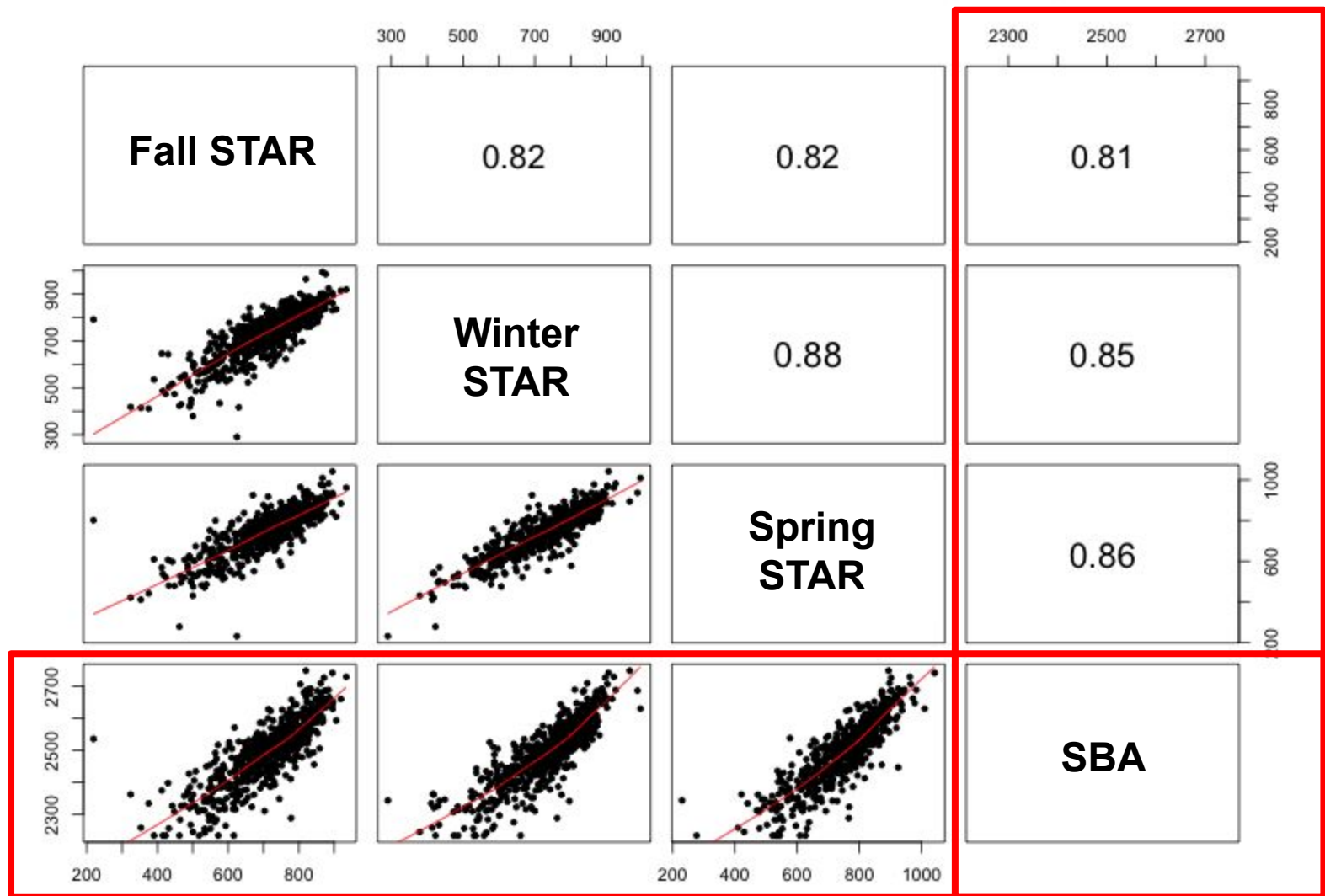
STAR Reading-SBA ELA Correlations - Grade 7 (N=459)



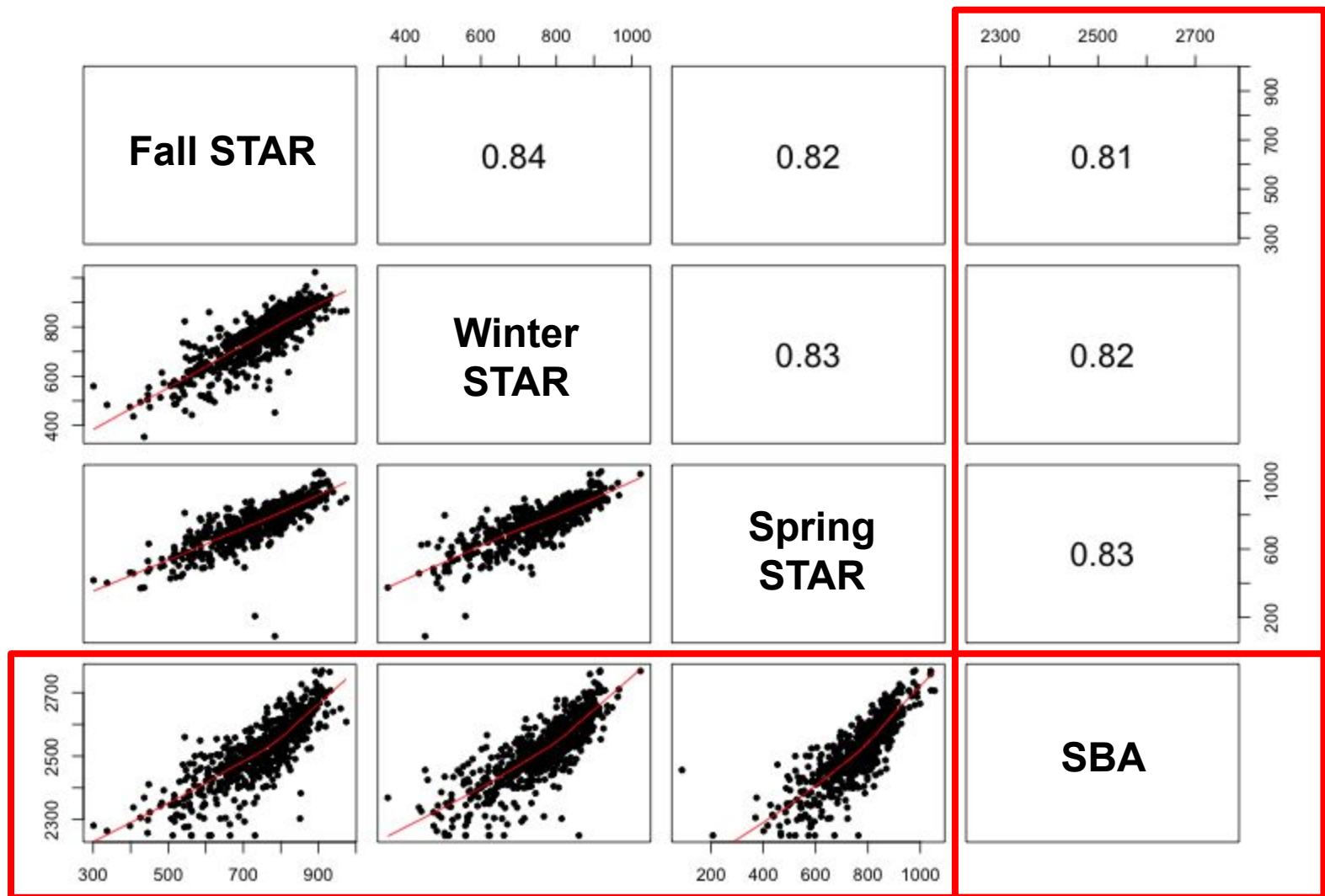
STAR Reading-SBA ELA Correlations - Grade 8 (N=417)



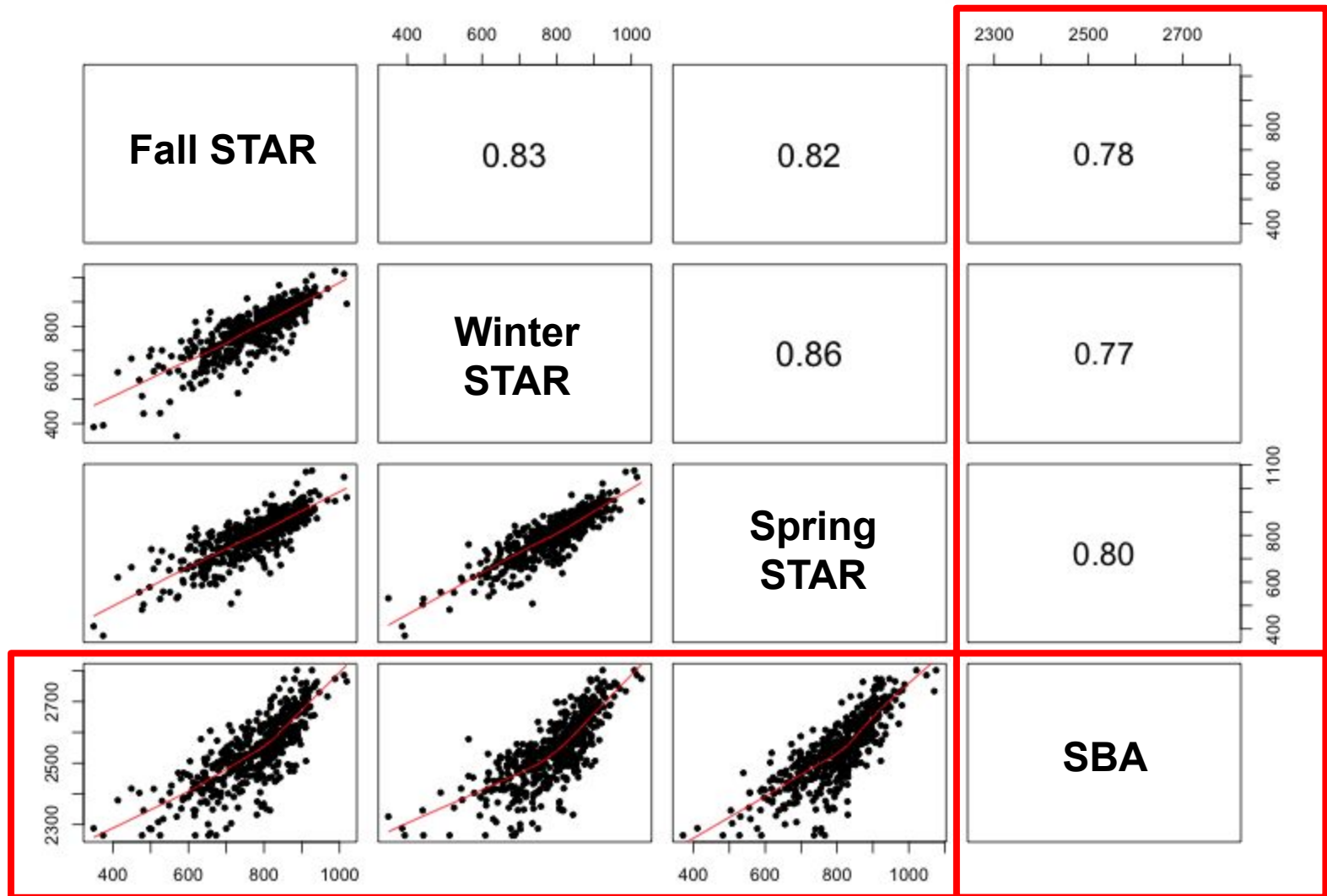
STAR Math-SBA Math Correlations - Grade 6 (N=613)



STAR Math-SBA Math Correlations - Grade 7 (N=621)



STAR Math-SBA Math Correlations - Grade 8 (N=465)



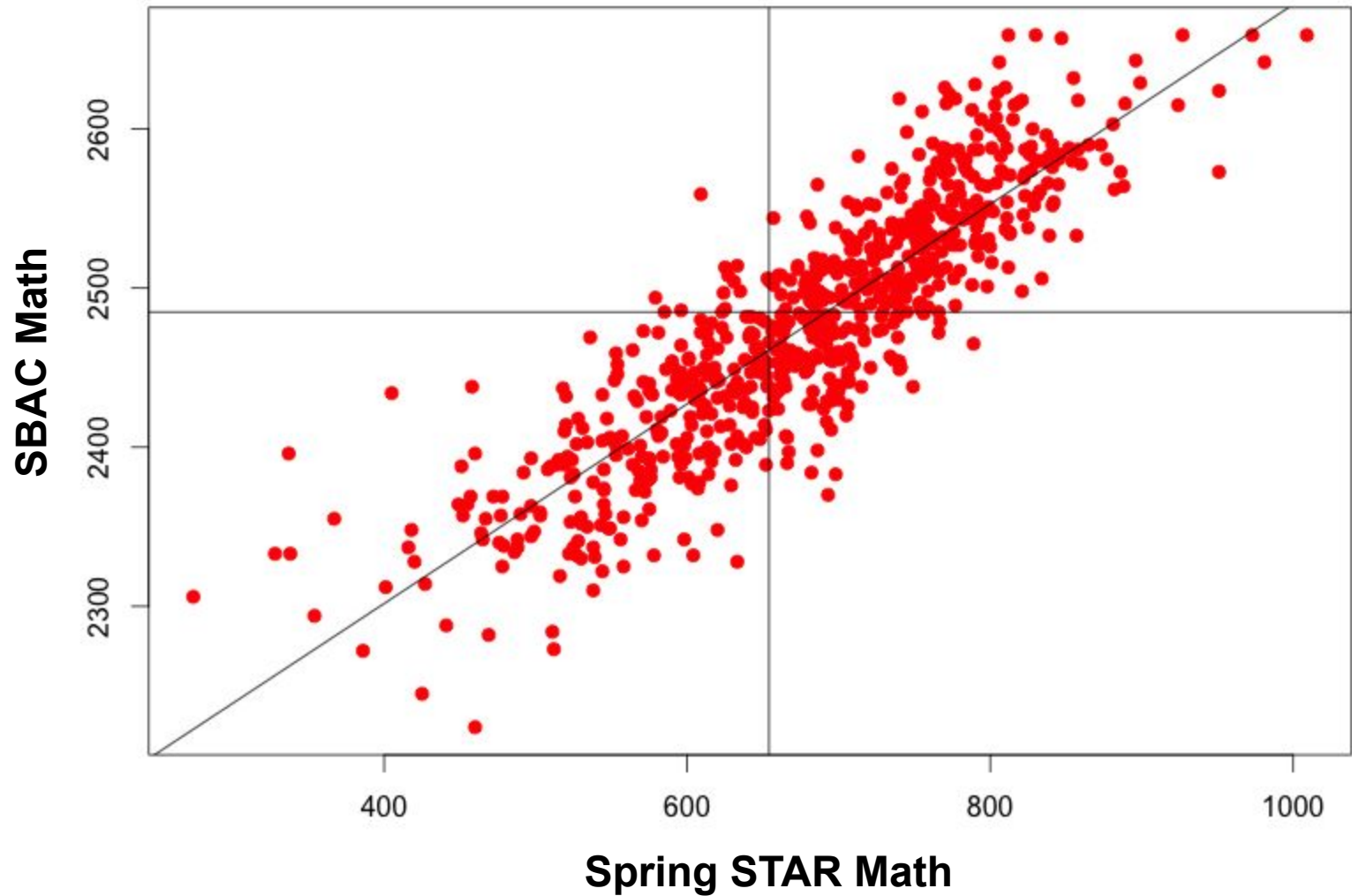
Results of Renaissance Linking Study¹

Grade	STAR Reading - SBA ELA		STAR Math - SBA Mathematics	
	N	r	N	r
3	11,068	0.82	10,800	0.84
4	12,200	0.82	10,582	0.86
5	10,908	0.83	9,750	0.86
6	8,072	0.81	7,852	0.86
7	6,320	0.83	6,344	0.86
8	6,209	0.83	5,424	0.83

¹Renaissance Learning. (2015, October 30.). *Relating STAR Reading and STAR Math to Washington Smarter Balanced Assessments performance.*

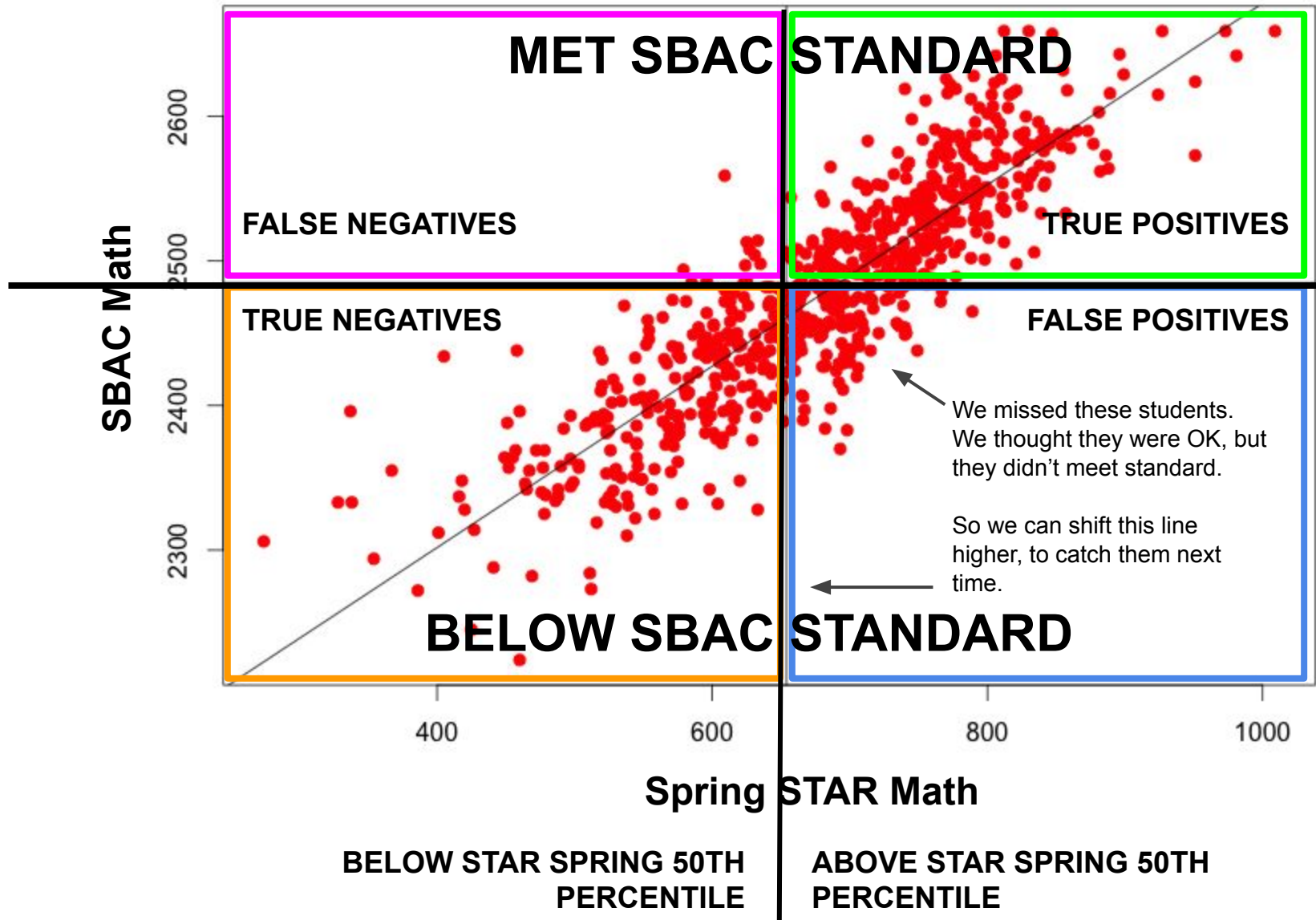
A Very Strong Correlation: $r = 0.87$

Grade 4 Math



A Very Strong Correlation: $r = 0.87$

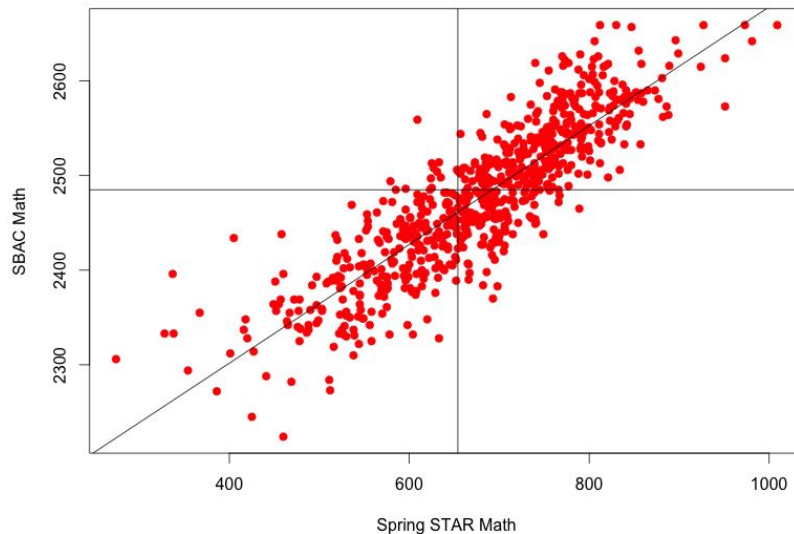
Grade 4 Math



Turn and talk

What are you thinking?

What does this information mean?



A reaction to Smarter Balanced testing...

To paraphrase . . .

“It’s [the Smarter Balanced assessment] really just measuring SES . . . because only kids of high SES could pass this thing.”



Validity evidence: Correlations and coefficient of determination

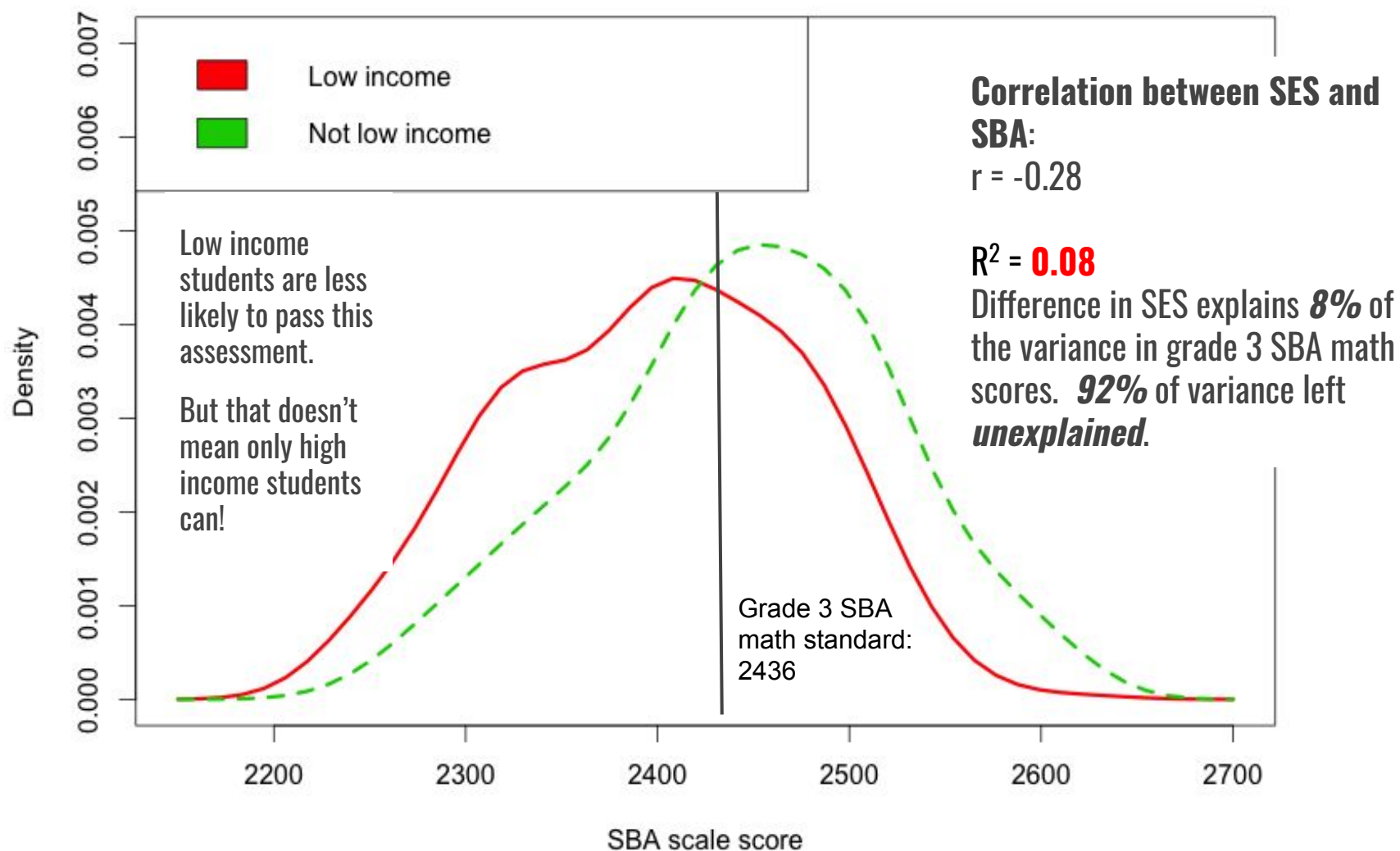
Indeed, differences in socioeconomic status will influence achievement on this assessment like any other

We can investigate this scientifically by looking at correlations (r) and the coefficient of determination (R^2)

Low values would indicate SBA is measuring far more than this economic contrast between students

Validity evidence: Effect of SES on SBA achievement

Grade 3 Mathematics



Validity evidence: Effect of SES on SBA achievement

Statistics by grade level

Grade	English Language Arts			Math		
	N	r	R ²	N	r	R ²
3	709	-0.29	0.08	703	-0.25	0.06
4	690	-0.24	0.06	706	-0.26	0.07
5	697	-0.20	0.04	660	-0.21	0.04
6	747	-0.18	0.03	747	-0.23	0.05
7	706	-0.21	0.04	706	-0.24	0.06
8	806	-0.25	0.06	806	-0.22	0.05

Correlational statistics across grade levels and content reflect the graph we saw on the previous slide. The correlation reflects a sizeable achievement gap, but that does not mean low income students cannot pass the assessment, or that only high income students can.

Lessons to inform our leadership moves

Can't go wrong being clear about intended uses of assessments

- Identify students at risk, trigger interventions, progress monitoring
- Formative assessment and resource system for teachers, PLCs
- Frame SMART goals for district and school improvement plans
- Measure and monitor change in achievement/growth gaps
- Evaluate instructional program K-12

Lessons to inform our leadership moves


Use data to address validity concerns, seize teachable moment

- Correlation data assuaged some doubts about validity of the STAR assessment, use of data
- Raised assessment literacy, clarified thinking about assessment
- Lowered barrier to full testing participation, use of data for SIP goals
- Stimulated interest in more professional development

Lessons to inform our leadership moves

Frame challenge as a chance to reflect on, refine our district assessment plan

- Are the data worth the cost of collecting them?
- Are we overtesting?
- Where are we heavy? Where are we light?
- What capacity do we have to reflect on the results?
- Are we data rich but analysis poor?
- What's the most efficient model to get everyone the data they need at the lowest cost?

A black and white photograph showing a close-up of a typewriter. A sheet of paper is inserted into the machine, and the word "Questions?" is typed in a classic typewriter font. The top of the keyboard is visible at the top of the frame, showing several keys. The paper has a slightly textured appearance, and the lighting creates soft shadows on the typewriter's body.

Questions?

THANK
YOU