OREGON
CUSD220
Academics | Activities | Service | Leadership

Date: Sep 17, 2018

From: Adam Larsen, Assistant Superintendent

To: Board of Education

Cc: Thomas Mahoney, Superintendent

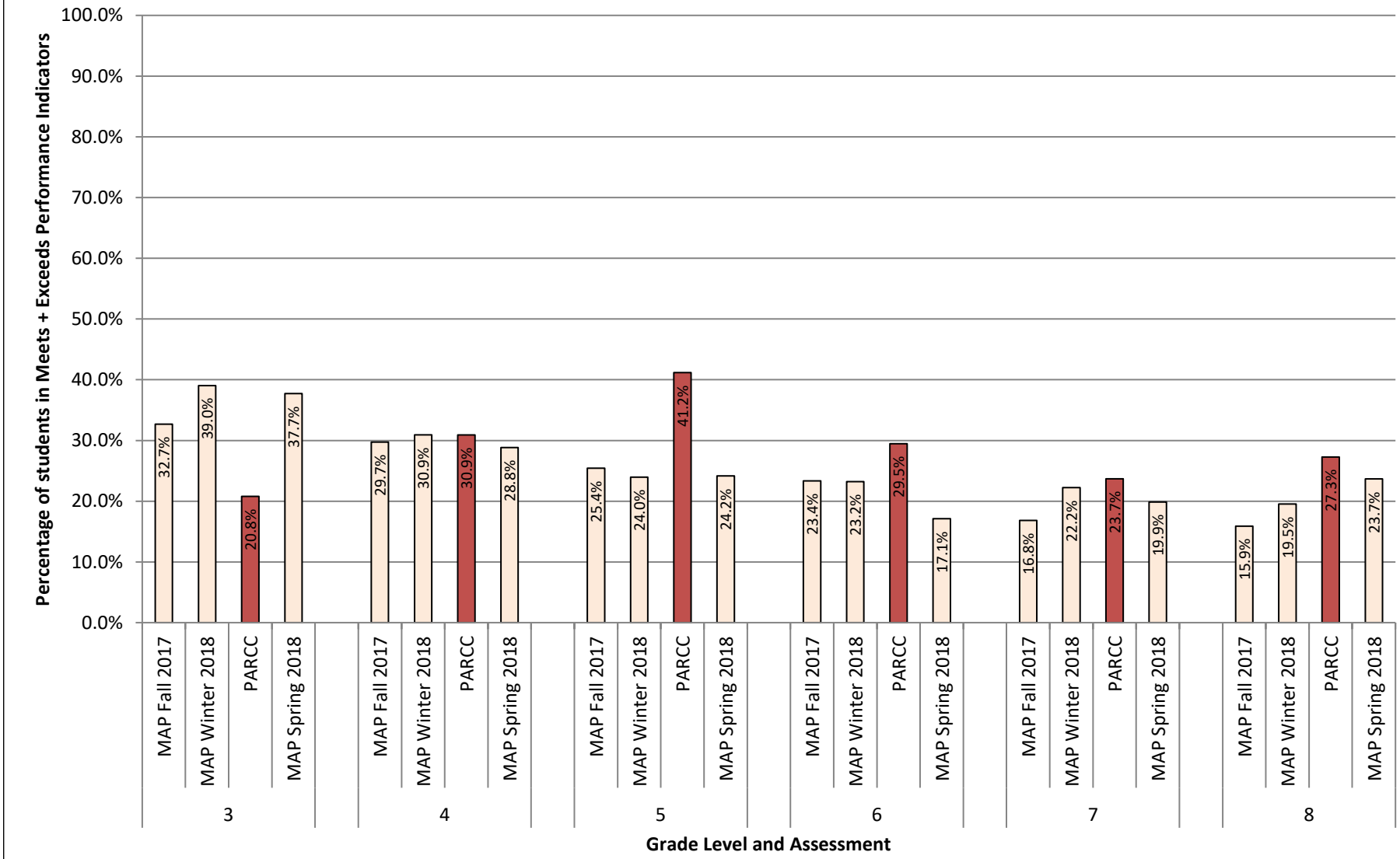
Re: September 2018 Board Report

2018 PARCC Data

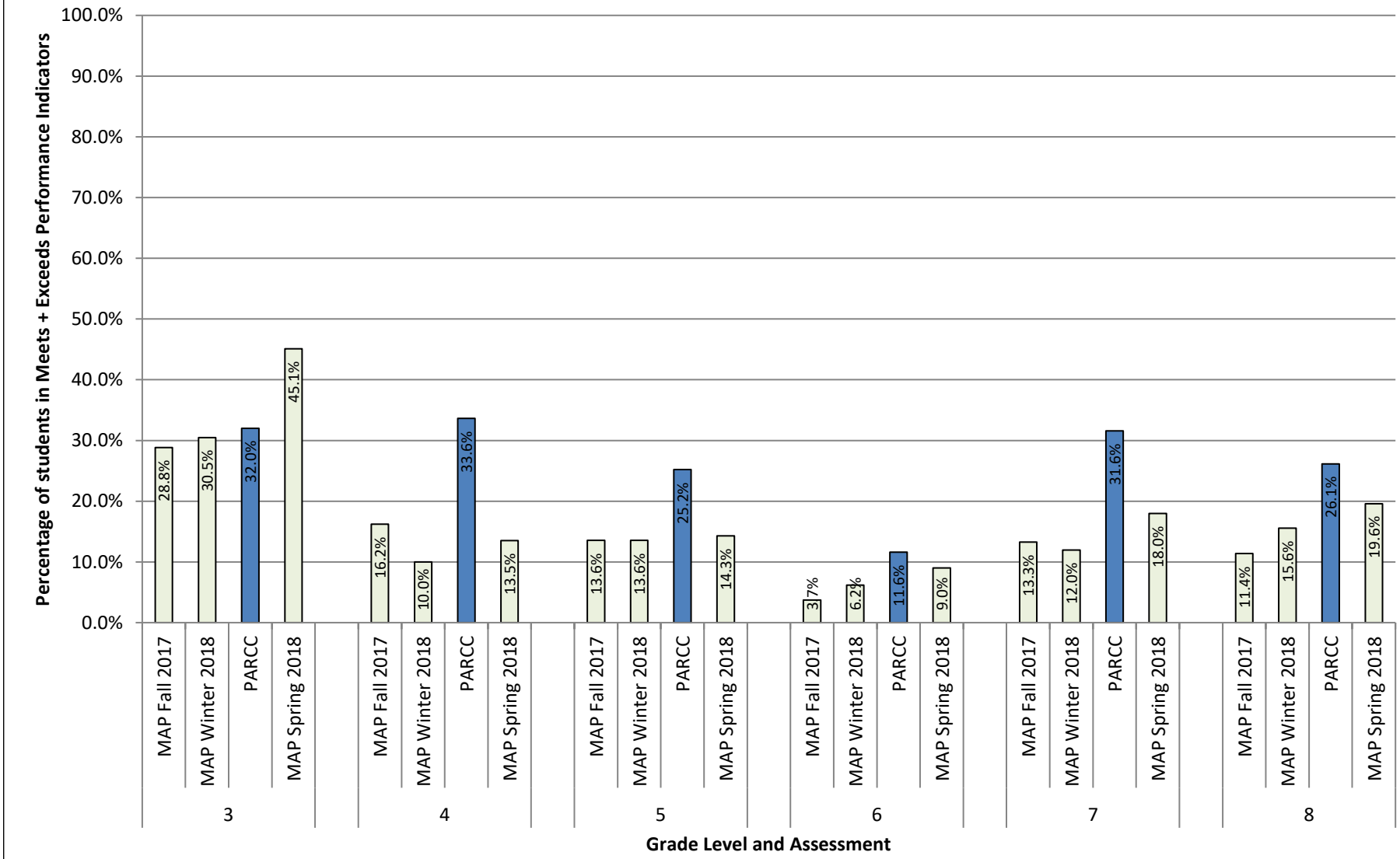
In Spring 2018, students completed the third annual administration of the Partnership for Assessment of Readiness for College and Careers (PARCC) test. Our implementation of the assessment went very smoothly, and we received scores in a timely fashion this summer. While final reports are currently being generated, we have some raw data and can begin to mine the data for trends.

One of the first analyses we usually conduct places NWEA Measures of Academic Progress (MAP) data against PARCC data to see how the percentages compare. We start by looking at the simple percentages by grade level and subject. That graph follows.

2018 PARCC Reading
with Fall, Winter, and Spring Predictions from MAP



2018 PARCC Mathematics
with Fall, Winter, and Spring Predictions from MAP



Because we spend a great deal of time predicting future scores for the purposes of intervention, we always compare accuracy of MAP-PARCC relationship at the individual student level as well. This is useful information in deciding whether to continue use the MAP test in the future.

The present approach involves identifying the types of possible errors and indicating their prevalence in the statistical sample. This analysis supposes that the default condition of a student is to meet expectations on the PARCC. This is referred to as the *null hypothesis*. For each student, the MAP test is used to identify students where the null hypothesis should be *rejected*, which would indicate that he or she will not meet expectations on PARCC. When a student is predicted to meet expectations on the PARCC, it is said that the null hypothesis *fails to be rejected*.

At the time of prediction (Fall MAP), there are two conditions in which a student may fall: predicted to meet expectations or predicted not to meet expectations. At the time of the final assessment (Spring PARCC), there are also two conditions: meeting expectations, and not meeting expectations. For simplicity, these conditions are referred to as *over* and *under* (short for *over the bar* and *under the bar*) going forward. When these two sets of conditions are crossed, a table such as below emerges:

OREGON
CUSD220
Academics | Activities | Service | Leadership

		Actual Performance (Spring 2018 PARCC)	
		Over	Under
Predicted Performance (Fall 2017 MAP)	Over	Correct	Type II Error
	Under	Type I Error	Correct

Predict: Over, Actual: Over – This student was correctly identified as **not** being at risk for falling below expectations on the PARCC. This is commonly referred to as a "correct miss."

Predict: Under, Actual: Under – This student was correctly identified as being at risk for falling below expectations on the PARCC. This is commonly referred to as a "correct hit."

Predict: Over, Actual: Under – This student was predicted to meet expectations on the PARCC, but fell below on the actual test. This type of incorrect prediction is known as a Type II error in research. In practice, this is a student who "fell through the cracks" of the intervention system. Because the student was not expected to score below the state-mandated benchmark, he or she was likely not targeted for additional intervention designed to remediate the skills in which the deficiencies lie. An alternative hypothesis is that the student had a bad test day when he or she took the PARCC.

In the school setting, a Type II Error is considered worse than Type I because the student was not identified as needing additional assistance when it probably would have helped.

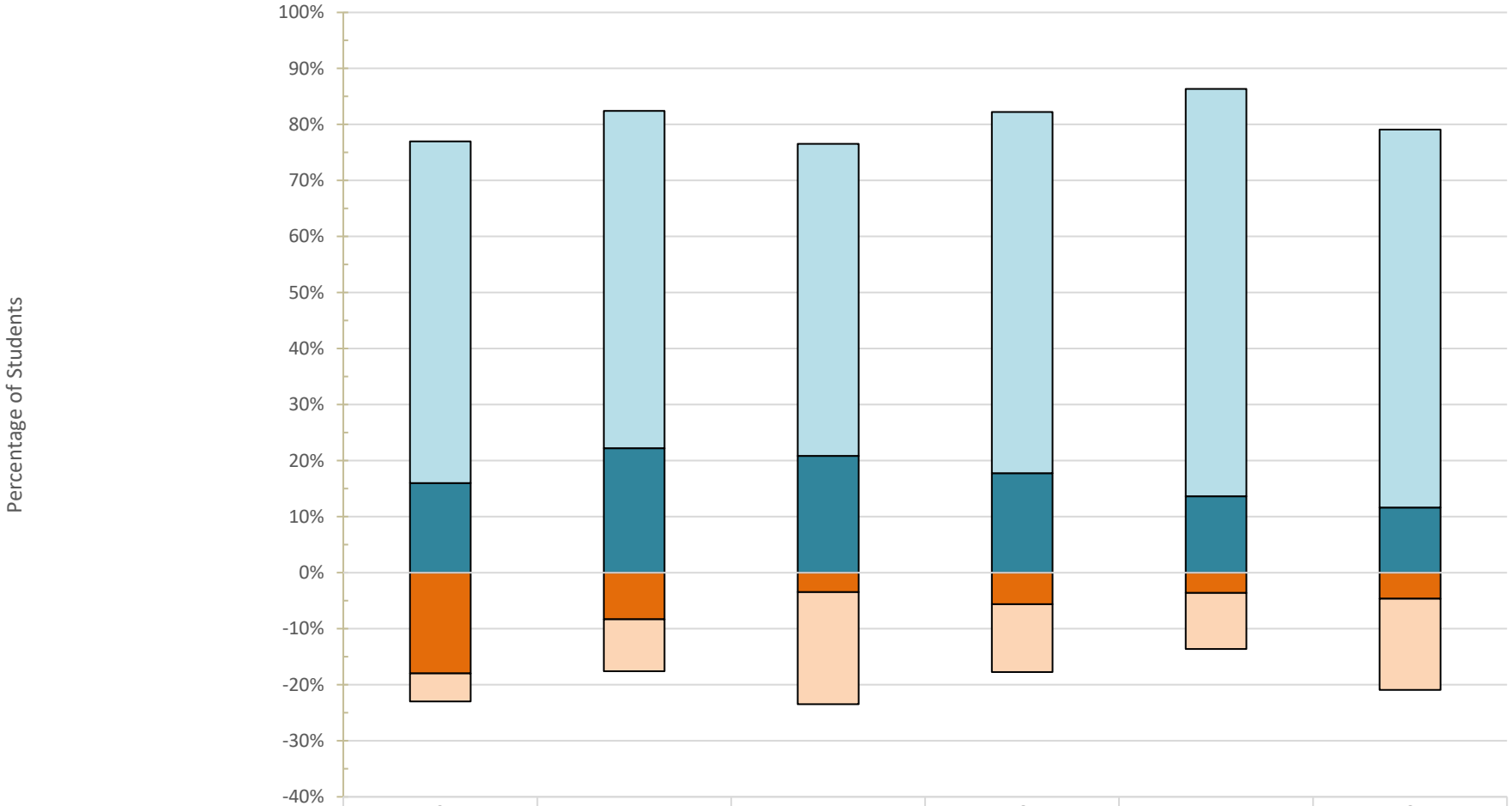
Predict: Under, Actual: Over – This student was predicted not to meet expectations on the PARCC, but performed successfully on the actual test. This type of incorrect prediction is known as a Type I error in research. In practice, this is a student who was targeted for intervention, and the intervention was successful in bringing that student up to expectations by the time of the PARCC. An alternative hypothesis is that the student had a bad test day when he or she took the MAP.

In the school setting, a Type I Error is considered more acceptable than a Type II error, because students on the bubble are being over identified for intervention. These students, while they did meet expectations on the PARCC, may have only done so because of the intervention in place.

The rates of these two types of errors are related. If high cutpoints are used for identification, then more Type I errors will be committed. Lowering the cutpoints results in a lowered Type I error rate, but a higher Type II error rate. The selected cutpoints strike a balance between these two.

The percentage of students falling into each of these four cells was computed and plotted for visual inspection:

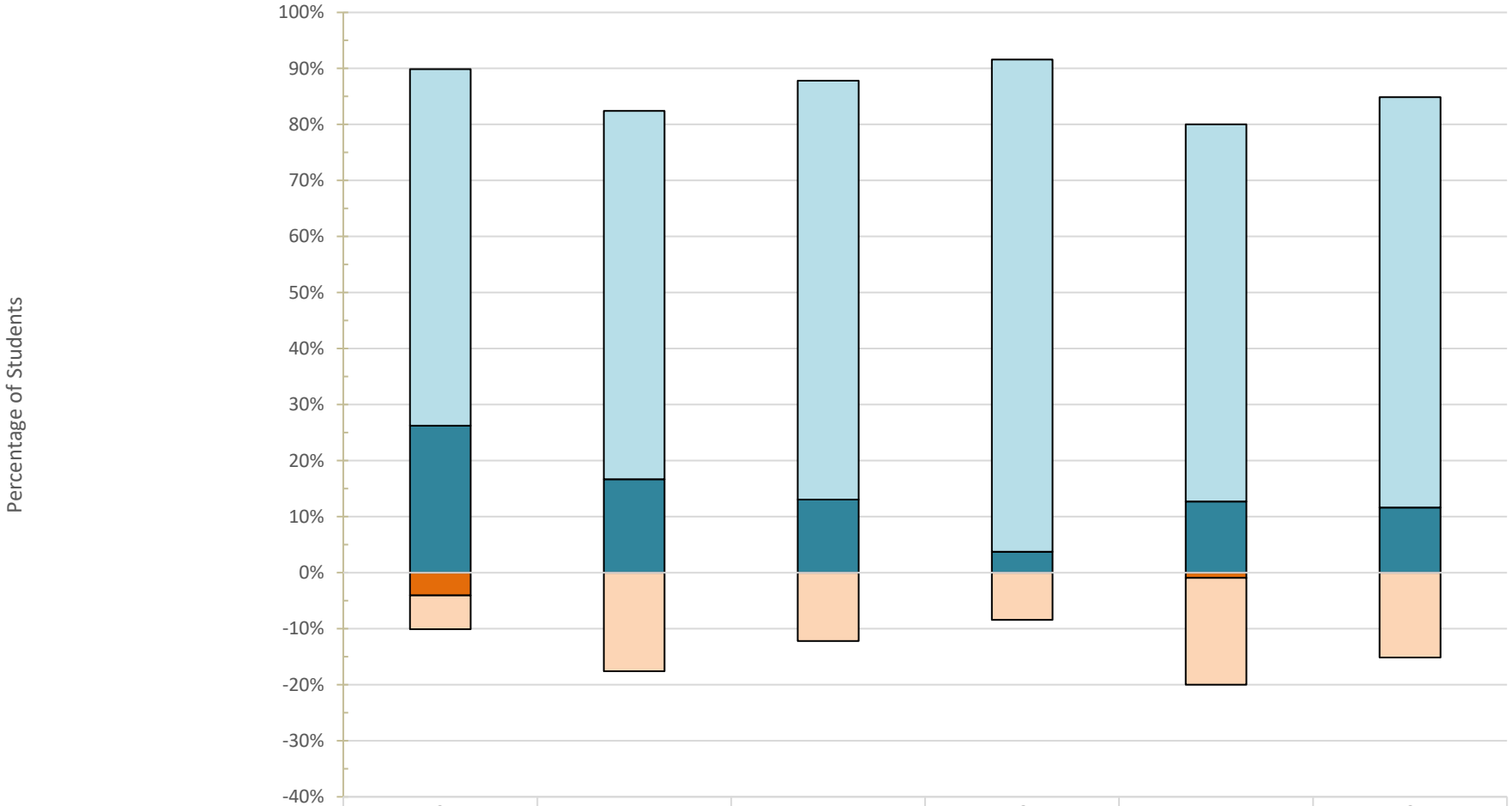
2017-2018 MAP-PARCC Prediction Accuracy
English/Language Arts



	3	4	5	6	7	8
Predict: Not Meet, Actual: Meet	-5%	-9%	-20%	-12%	-10%	-16%
Predict: Meet, Actual: Not Meet	-18%	-8%	-3%	-6%	-4%	-5%
Predict: Not Meet, Actual: Not Meet	61%	60%	56%	64%	73%	67%
Predict: Meet, Actual: Meet	16%	22%	21%	18%	14%	12%

Grade Level

2017-2018 MAP-PARCC Prediction Accuracy
Mathematics



	3	4	5	6	7	8
Predict: Not Meet, Actual: Meet	-6%	-18%	-12%	-8%	-19%	-15%
Predict: Meet, Actual: Not Meet	-4%	0%	0%	0%	-1%	0%
Predict: Not Meet, Actual: Not Meet	64%	66%	75%	88%	67%	73%
Predict: Meet, Actual: Meet	26%	17%	13%	4%	13%	12%

Grade Level

OREGON
CUSD220
Academics | Activities | Service | Leadership

Conclusions

MAP continues to be a reliable predictor of future performance on the PARCC. Students projected to meet the state-mandated cutscore on the spring test almost always score at or above this level. When the spring measure was the ISAT, we typically saw more Type I errors than Type II. This was encouraging, as it suggested to us that intervention efforts were effective, and more students rose to the assessment challenge, rather than fell through the cracks as the year progressed.

This is no longer the case with PARCC. We are seeing more students whose fall scores suggest that they will meet or exceed in the spring, but then fail to perform at that level. Further study will be required to determine whether this is a result of actual regression during the school year or whether the MAP cutscores are not rigorous enough to predict PARCC scores with the same accuracy.

OREGON
CUSD220
Academics | Activities | Service | Leadership

Remark Case Study

We have used a pieces of software called Office OMR for a number of years and for a variety of purposes. The program allows us to print our own Scantron-like "bubble sheets" to collect data from students and staff. Most frequently, we use this for survey instruments where online tools are impractical. Our original purpose for buying the software was so we could administer practice ACT exams to students. This would allow us to collect data about each question and question type, as well as provide a real-looking score report to each student.

When Illinois transitioned away from the ACT, we moved our test preparedness processes to the SAT Suite of assessments (PSAT8/9, PSAT10, and SAT). For the past two years, we have only given one practice SAT to juniors in February, but beginning in 2018-2019, we will administer assessments according to the following schedule:

Source	Test	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Remark	Practice PSAT9	x									
State	PSAT9									x	
State	PSAT10									x	
Remark	Practice SAT (11)			x				x			
State	SAT (11)									x	

This framework will help us check in periodically with students on their mastery of the skills that the SAT targets. Teams are forming to review the data and create intervention strategies to help the students improve their performance.

When we gave students a practice SAT in February, I tweeted out a series of photos and captions about our process to share with my colleagues in educational data science. Remark actually reached out and asked about doing a case study about how we use their software. We had a couple of phone conferences to lay out the study and edit the copy, and they shared it on their blog over the summer. This was pretty neat to be featured by a company whose software is used all around the world.

This is the direct URL of that blog post, but I have also attached the pages below.
https://remarksoftware.com/case-studies/2018/07/illinois-district-preps-students-for-the-sats/

Illinois District Preps Students for the SATs

Background

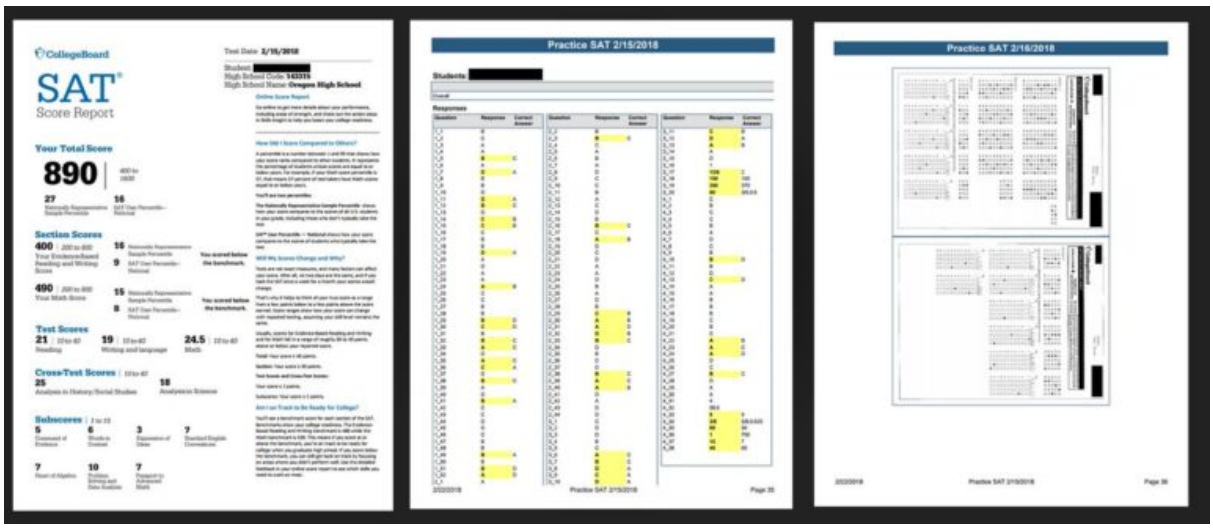
Adam Larsen is the assistant superintendent of an Illinois school district. He and the rest of the staff are dedicated to helping their students do well on the College Board Organization's Scholastic Aptitude Test (SAT®). During a School Improvement Day, high school students took an official **College Board** released practice test, while simulating actual testing conditions. The staff used **Remark Office OMR** to analyze the best way to prep their students for the real exam.

The Problem

The SATs can be very stressful for students and educators alike. In order to ensure that his students do well, Mr. Larsen wanted to know how best to teach them the essential information for the test, but it was unclear what exactly his students did and did not know.

The Solution

The staff first created its own bubble sheet for the practice test using Microsoft Word, and graded the responses with Remark Office OMR. Within a day or two, teachers could distribute an authentic looking score report, created in Microsoft Word, as well as students' graded responses and answer explanations.



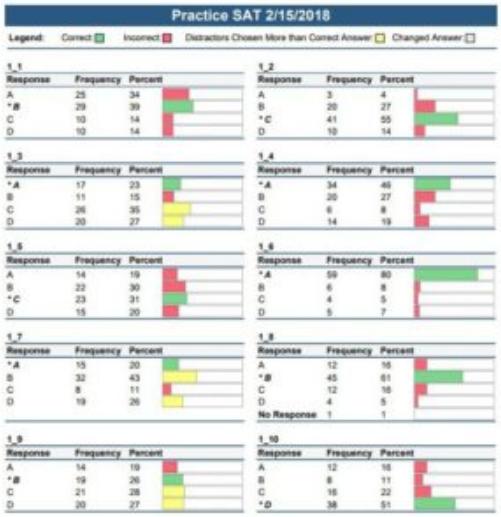
With Remark Office OMR and Microsoft Word, students can receive a genuine-looking score report, to make the experience seem more real, along with helpful information about questions they missed.

The teachers then used different report functions from the Remark software to obtain a thorough understanding of what questions were missed and why. For example, they ran a report that found the most frequently missed and correctly answered questions, along with an analysis of the reliability using point-biserials and p-values.

They also found common misconceptions in word definitions and formulas with frequency distributions of response patterns for each question. The questions in which distractors were chosen more often than the correct answers contained the misconceptions.

The teachers then combined the statistical data about student responses with the test questions to eliminate the need to go back and forth from the test document to the reports.

Finally, all of the raw score data was exported to Microsoft Excel and weighted accordingly to find each student's "SAT Score". Teachers could then use the spreadsheet to see each student's overall score as well as sub scores in math, reading, and writing. In addition, sub topics like "Words in Context" and "Problem Solving and Data Analysis" were also analyzed and reported.



Frequency distribution analysis shows that students chose two distractors more often than the correct answer on this question.

The Conclusion

With all of their newly collected data, teachers at the school are forming different specialized groups for students to help with the progression of newly identified areas of improvement. Mr. Larsen himself noted that the outcome of this experience was, **"a closely-simulated test experience, timely corrective feedback, focused discussions about student performance, and deliberative practice".**

Try **Remark Office OMR** out for yourself with the **free demo version** or **sign up for a webinar** to see it in action!

Demo Webinar

Mailing Address

17 General Warren Blvd.

Contact Information

Remark Sales

OREGON
CUSD220
Academics | Activities | Service | Leadership

Dissertation Process

I was notified on August 10 that I officially passed the candidacy examination, which is a part of the dissertation process at NIU. This was an eight-week writing assignment based on my research interests and intended to guide further inquiry about my potential research questions. The professors provided valuable feedback on my writing and research ideas, and I am well-poised to move into the proposal phase.

During this next step, we will be writing a tentative first three chapters of the dissertation. This includes the introduction, literature review, and methodology sections. At present, the plan is to implement an early warning system similar to what we have here, but at a different school. This would involve developing a focus group protocol for identifying important risk factors, performing the statistical analysis to develop their model of risk, then training staff and implementing the tool at that school. The project has a green light from the committee but still needs to be proposed and defended, and then cleared by the Institutional Review Board. I will continue to provide updates as this process moves forward.

Respectfully Submitted,

Adam P. Larsen
Assistant Superintendent
Oregon CUSD #220