

education  
analytics



# *What Are We Missing?: The Impact of COVID-19 on Educational Data Quality and Availability*

June 2021

Presented by Selina Eadie Dhillon

# *What this presentation will cover:*

During today's video presentation, we will discuss:

1

The impact of COVID-19 and school closures on data availability and data quality

2

EA's process for assessing tradeoffs and implications of missing data on analytics and reporting

3

An example of how EA has grappled with this issue using our work on the CORE dashboard as a case study

# About Education Analytics (EA)

## About EA

---

- Non-profit based in Madison, WI with team members located all over the country
- Work with over **72 partners** across **28 states**
- Our team includes data analysts, policy strategists, researchers, and education experts

## What we do

---

- Conduct research and develop rigorous analytics that support actionable solutions and continuous improvement in American education
- Help our partner make better decisions on policies and programs that lead to success for schools and all students

# The missing data challenge

- We all know COVID-19 and school closures impacted data quality and availability

- What does school-day attendance look like?
- How does this impact chronic absenteeism?
- How is gradebook data affected?

Virtual  
instruction



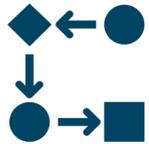
- In the absence of test data, how do we measure learning?
- How do we measure students' year-over-year growth?

Cancelled state  
tests



What academic metrics are still possible to calculate, and how do we determine what we can and cannot responsibly report out to stakeholders?

# The different ways data quality can impact education metrics and reporting



## Data warehousing / engineering and metric calculations

- How are non-COVID year business rules impacted by data quality?
- How do we process files with the same structures but missing values?



## Modeling

- Do metrics mean the same things they did during non-COVID years?
- With what caution should we interpret findings?



## Visualization and reporting

- How should we display metrics with missing or incomplete data?
- How do we transparently report limitations?



## Supporting use

- How do we work with our partners to understand how they plan to use COVID-impacted metrics?
- How do we help them communicate about impact to their stakeholders?

# How has EA approached missing data in our work?



We adapted our Missing Data Guide developed for our own work that you can use and apply to your own context

1. Reviewed all EA projects and identified those affected by COVID-year data
2. Identified key questions and best practices for examining data quality checks (QC)
3. Developed a Missing Data Guide that documented considerations for:
  - a. Data QC and metric calculations
  - b. Research methodology
  - c. User-facing dashboards

# *Examples of impacts that we saw:*

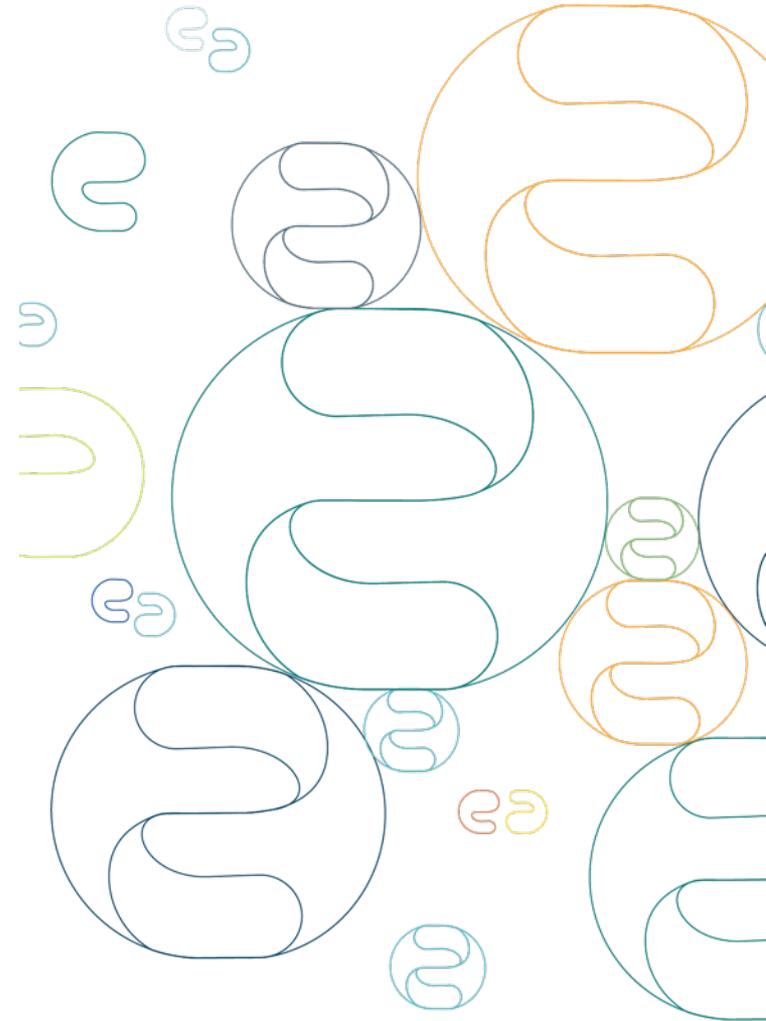
---

- Some schools' grading practices and required **business rule changes**
- Some **metrics** and student group identifiers were dependent on a fixed number of days in a school year, and needed to be adjusted
- Some **reports and dashboards** had to be redesigned or appended to acknowledge missing testing and growth data

# At a high level...

---

- We recognized that data quality impacts could be far-reaching and not immediately obvious to our partners
- **Next**, we'll walk through an example of a metric on a partner dashboard that was impacted by incomplete data and the decisions we made to address these data issues



# *CORE Dashboard Case Study*

# Case study: CORE Dashboard



## CORE Districts

- CORE is collaborative of school districts working together to improve student achievement through partnership

## Data System

- CORE has developed a multiple measures accountability system and worked with EA to build its underlying data system

## Dashboarding

- CORE gives schools and districts access to dashboard that includes achievement, academic growth, graduation, chronic absenteeism, EL re-designation rates, SEL data, and school climate survey data

# About the CORE dashboard

---

- The CORE dashboard features data covering ~4 million students across ~200 districts in the state of California
- To support the CORE dashboard, EA operates an **automated database system** to load, clean, organize, calculate, analyze, and present key indicators on an accessible dashboard
- Ensuring this system's **data quality** is our highest priority

# Case study: CORE Dashboard



CORE Index

LCAP

Network: CORE | District: Network Overall | School Level: ES | School Type: | School: | School Year: 2019-2020

## Index Results: Social-Emotional & Culture-Climate Domain (All Students)

	Metric Result 2019	Metric Result 2020	Change in Metric Performance from 2019 to 2020	Index Level 2020	Change in Index Level from 2019 to 2020
Chronic Absenteeism *	16.7% <small>Chronically Absent 2019</small>	10.3% <small>Chronically Absent 2020</small>	-6.4%	5 out of 10	↑ 2
Suspension Rates (includes students suspended and/or expelled) *	0.9% <small>Suspended (and/or Expelled) 2019</small>	0.6% <small>Suspended (and/or Expelled) 2020</small>	-0.3%	8 out of 10	↑ 1
Culture and Climate: FAMILY Overall	92.9% <small>Percent Favorable 2019</small>	94% <small>Percent Favorable 2020</small>	1.1%	7 out of 10	No Data
Culture and Climate: STAFF Overall	82% <small>Percent Favorable 2019</small>	81% <small>Percent Favorable 2020</small>	-1.0%	4 out of 10	→ 0
Culture and Climate: STUDENT Overall	76% <small>Percent Favorable 2019</small>	74% <small>Percent Favorable 2020</small>	-2.0%	3 out of 10	↓ 1
Social-Emotional Skills: Growth Mindset	80% <small>Percent Positive 2019</small>	72% <small>Percent Positive 2020</small>	-8.0%	3 out of 10	↓ 3
Social-Emotional Skills: Self-Efficacy	71% <small>Percent Positive 2019</small>	59% <small>Percent Positive 2020</small>	-12.0%	5 out of 10	↓ 4
Social-Emotional Skills: Self-Management	72% <small>Percent Positive 2019</small>	65% <small>Percent Positive 2020</small>	-7.0%	2 out of 10	↓ 4
Social-Emotional Skills: Social Awareness	72% <small>Percent Positive 2019</small>	67% <small>Percent Positive 2020</small>	-5.0%	2 out of 10	↓ 3

■ Above Average (Index Levels 8, 9 and 10)
 ■ Average (Index Levels 4, 5, 6, 7)
 ■ Below Average (Index Levels 1, 2, 3)

# Case study: Chronic Absenteeism

- Today, we will dive into the **chronic absenteeism metric** on the CORE dashboard

## CDE definition of Chronic Absenteeism

A student is defined as chronically absent if they are absent at least 10% of the instructional days in which they were enrolled at the school.

*Example: 20 of 2,000 students at Nebula Middle School were absent at least 10% of the instructional days. The chronic absenteeism rate for the school is 1%:  $20 \text{ divided by } 2,000 = 1\%$*

# Three domains considered

---

- For the CORE dashboard, we dove into three domains to understand how chronic absenteeism was impacted

Data quality/  
metric  
calculations

Research &  
methodology  
decisions

Front end-  
visualizations

---

# *Data Quality & Metric Calculations*

# Three domains considered

---

- For the CORE dashboard, we dove into three domains to understand how chronic absenteeism was impacted

Data quality/  
metric  
calculations

Research &  
methodology  
decisions

Front end-  
visualizations

# Three domains considered

---

- For the CORE dashboard, we dove into three domains to understand how chronic absenteeism was impacted

Data quality/  
metric  
calculations

Research &  
methodology  
decisions

Front end-  
visualizations

# *How did COVID year missing data affect chronic absenteeism?*

---

After performing **quality control checks** on attendance and chronic absenteeism data, we saw...

- Methods of measuring attendance were different across districts given the limitations of the remote environment
- Most districts only reported attendance through **mid-March**, but some reported full-year attendance

To address this flag, we looked at the correlation between partial year and full year chronic absenteeism rates. Our “research and methodology” section will discuss our findings!

# How did COVID year missing data affect chronic absenteeism?

**Note:** We use continuous enrollment to link metrics to student demographic groups, so this adjustment was necessary for us to display how different student groups perform!

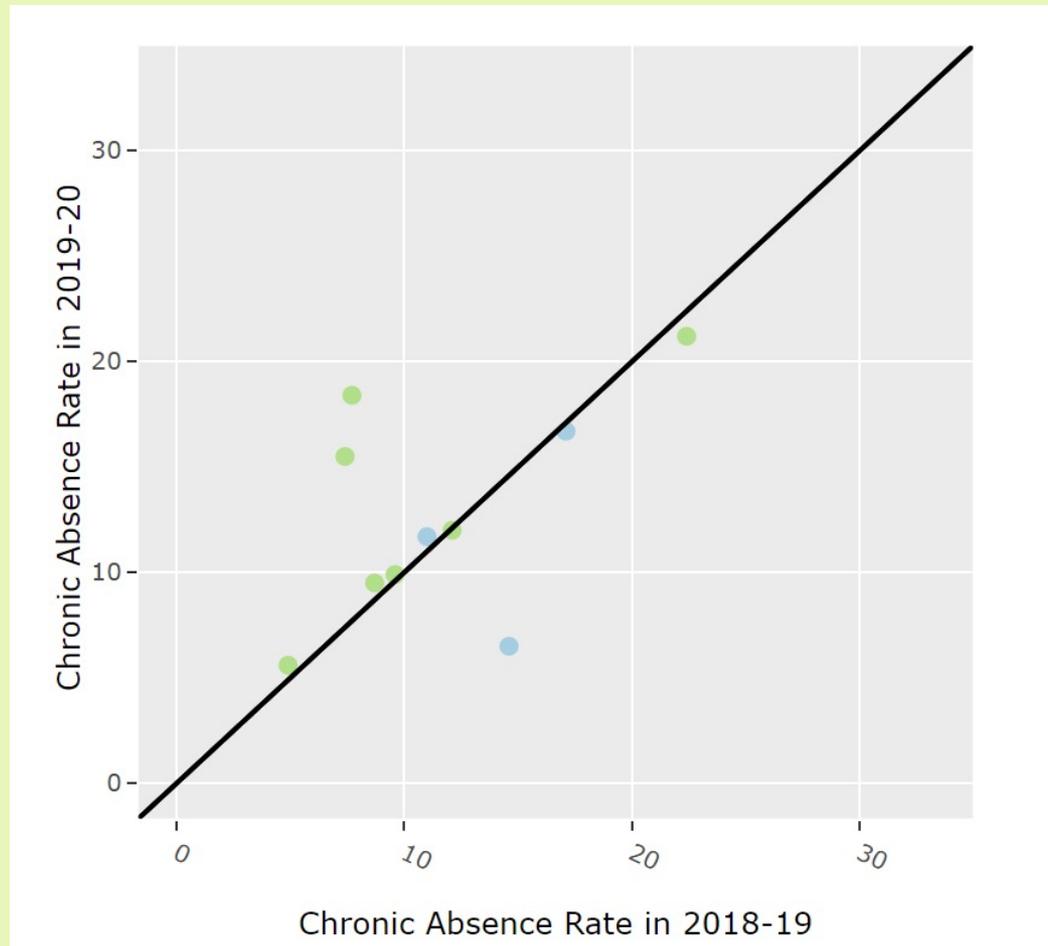
After reviewing the **business rules** and **file sources** for calculating chronic absenteeism, we learned...

- We needed a new source of truth for the end date for continuous enrollment to link chronic absenteeism to student demographic groups



**Decision:** We set February 29th as the 'cutoff' date for end of school year, or the last day of inclusion, for continuous enrollment (based on CDE guidance)

## Using visualizations to review data: District-facing reports



Schools % chronically absent in 2018-19 (x-axis) and 2019-20 (y-axis)



# Missing data questions for your team to ask about data quality & metric calculations

## Data and metric questions

- What are all the source files you expect to be missing or incomplete?
- Do you use any year-dependent parameters that need to be updated?

## Data quality and data equity checks

- Are there any additional data quality control checks to perform that might not be run in a normal year?
- **Equity check:** Are certain data limitations specific to certain student groups?\*

## Year-over-year review

- Are there any anomalies in year-over-year rates?
- Are there any differences in student sizes across years?

\* In CA, English Learner students were unable to take the ELPAC assessment in Spring 2020, which is used to determine English Learner Reclassification

---

# *Modeling and Methodology*

# Three domains considered

---

- For the CORE dashboard, we dove into three domains to understand how chronic absenteeism was impacted

Data quality/  
metric  
calculations

Research &  
methodology  
decisions

Front end-  
visualizations

# Three domains considered

---

- For the CORE dashboard, we dove into three domains to understand how chronic absenteeism was impacted

Data quality/  
metric  
calculations

Research &  
methodology  
decisions

Front end-  
visualizations

# About the CORE dashboard: Modeling and methodology

---

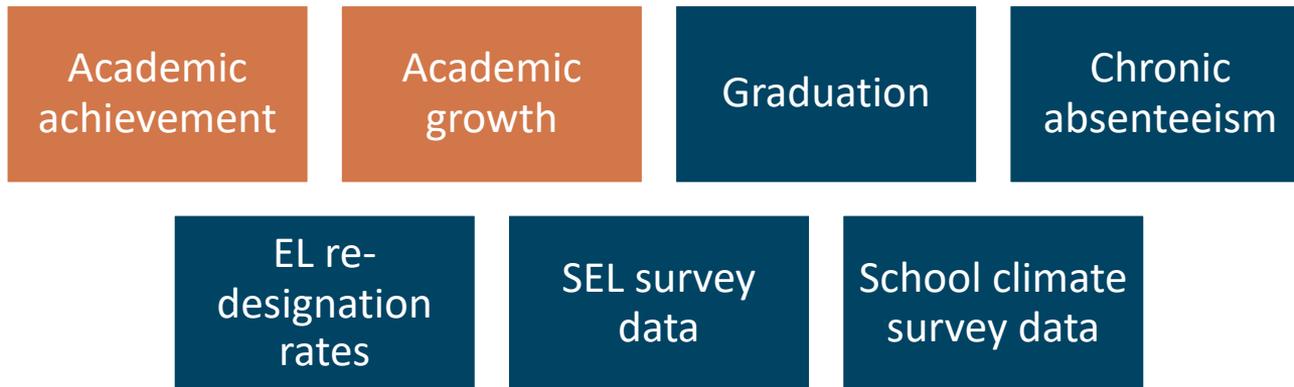
- To understand the impact of COVID on statistical models, the different metrics required different levels of methodology review based on their complexity and data sources/data availability



# About the CORE dashboard: Modeling and methodology dashboard

---

- For example, achievement and growth could not be calculated for SY 19-20 due to lack of assessment data



# Calculating growth with missing 2019-20 data

---

- Research team identified two options for growth measures:

1

Predict missing assessment outcomes, which requires multiple strong methodological assumptions and sets of simulations to verify them

2

Measure growth over a two-year period, from SY 2018-19 to SY 2020-21

# Calculating growth with missing 2019-20 data

- Given the technical complexity of the first option and the state's policy decision to waive mandated tests in 2019-20, CORE and EA jointly decided to support the two-year growth option
  - Achievement metrics were not reported on the dashboard for 2019-20

1

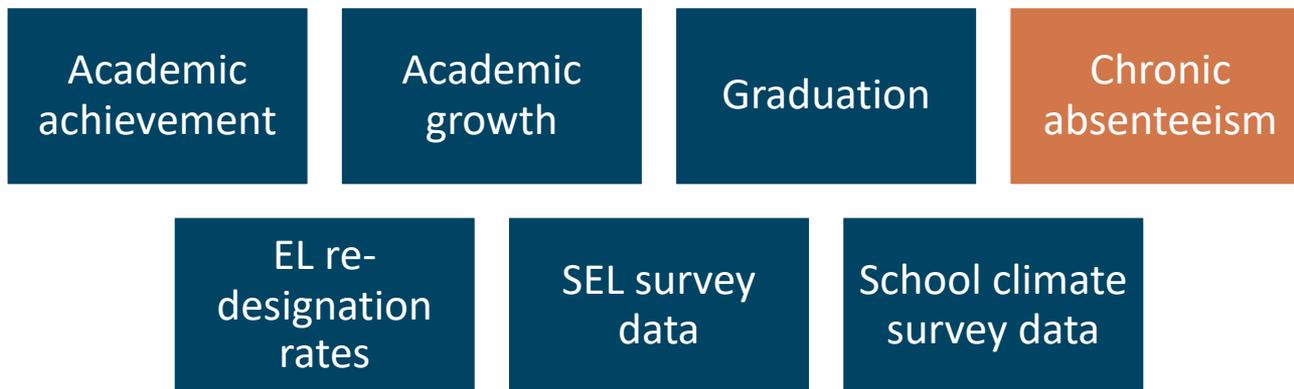
Predict missing assessment outcomes, which require multiple strong methodological assumptions and sets of simulations to verify them

2

Measure growth over a two-year period, from SY 2018-19 to SY 2020-21

# How did COVID year missing data affect chronic absenteeism?

- Chronic absenteeism could be calculated, because in most cases we had attendance data reported through March 2020
- Our question: With what **caution** should we **interpret findings**?



# *How did COVID year missing data affect chronic absenteeism?*

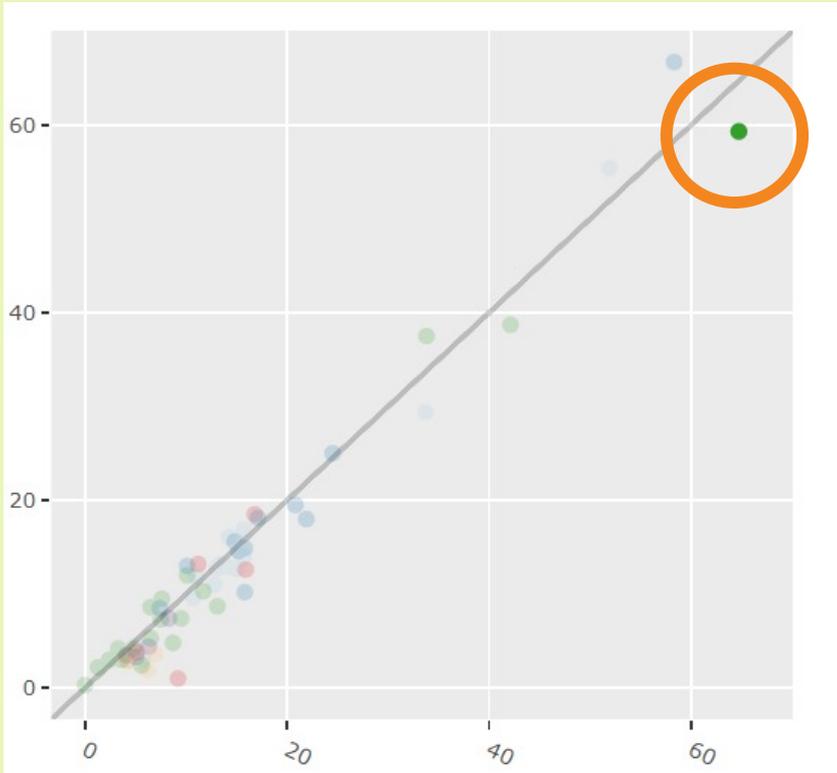
---

- Next step: compare attendance and chronic absenteeism rates from previous years to the 2019-20 impacted year
- Measure rates for chronic absenteeism **were well-correlated** between the 18-19 and 19-20 school years across the data collaborative
  - For a subset of schools: 2019-20 data was not a good proxy for full-year data

Every dot represents a school, and the color represents its district.  
Hover to see both school and district information.

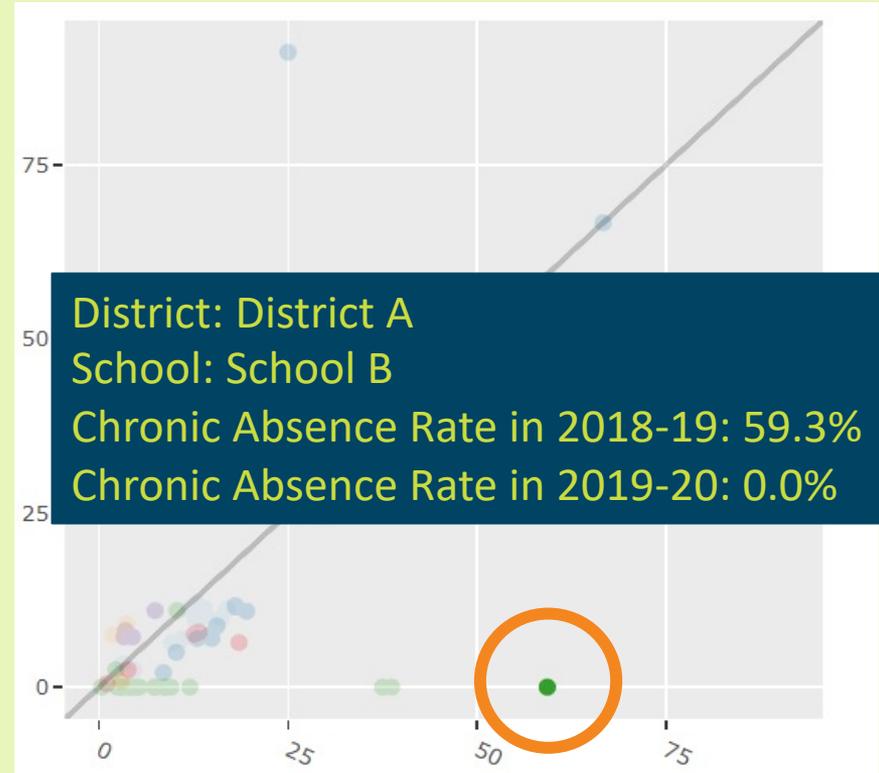
## Compare Chronic Absence Rates Across Years at School-Level

Chronic Absence Rate in 2018-19



Chronic Absence Rate in 2017-18

Chronic Absence Rate in 2019-20



Chronic Absence Rate in 2018-19

# *How did COVID year missing data affect chronic absenteeism? (cont.)*

---

- Overall, the year-year-correlations were high
- Important to note that for 2019-20, this data was not used by CORE members for high-stakes decision making, due to the CDE's decision to suspend the state accountability dashboard



**Decision:** Due to the year-over-year correlation of chronic absenteeism across the data collaborative AND the lower stakes use of the data, CORE & EA decided to report chronic absenteeism for SY 19-20



# Missing data questions for your team to ask about research and methodology

1

How do we address the absence of missing standardized 2019/20 testing data with respect to our growth metrics / methodology?

2

What is the extent to which the 2019-20 data is comparable to the full-year data?

3

What is the extent to which metrics using the partial-year data have different interpretations compared with those using the full-year data?

---

# *Frontend Visualizations & Reporting*

# Three domains considered

---

- For the CORE dashboard, we dove into three domains to understand how chronic absenteeism was impacted

Data quality/  
metric  
calculations

Research &  
methodology  
decisions

Front end-  
visualizations

# Three domains considered

---

- For the CORE dashboard, we dove into three domains to understand how chronic absenteeism was impacted

Data quality/  
metric  
calculations

Research &  
methodology  
decisions

Front end-  
visualizations

# About the CORE dashboard: Background and COVID year implications

- CORE dashboard reports metric performance at a district and school level (and by student demographic group)
- Had to determine what metrics to show, hide, or annotate due to limited and missing data

Metric	Metric Result 2019	Metric Result 2020	Change in Metric Performance from 2019 to 2020	Index Level 2020	Change in Index Level from 2019 to 2020
Chronically Absent	16.7%	10.3%	-6.4%	5	↑ 2
Suspended (and/or Expelled)	0.9%	0.6%	-0.3%	8	↑ 1
Percent Favorable	92.9%	94%	1.1%	7	No Data
Percent Favorable	82%	81%	-1.0%	4	→ 0
Percent Positive	76%	74%	-2.0%	3	↓ 1
Percent Positive	80%	72%	-8.0%	3	↓ 3
Percent Positive	71%	59%	-12.0%	5	↓ 4
Percent Positive	72%	65%	-7.0%	2	↓ 4
Percent Positive	72%	67%	-5.0%	2	↓ 3

■ Average (Index Levels 4, 5, 6, 7)  
■ Below Average (Index Levels 1, 2, 3)

# District-Level Results: 2016-2017

	Metric Result 2016	Metric Result 2017	Change in Metric Performance from 2016 to 2017	Index Level 2017	Change in Index Level from 2016 to 2017
Math	<b>51%</b> Meet or Exceed Standards 2016	<b>55%</b> Meet or Exceed Standards 2017	<b>4.0%</b>	<b>6</b> out of 10	<b>↑</b> 1
Language	<b>22%</b> Growth Percentile 2016	<b>33%</b> Growth Percentile 2017	<b>11.0%</b>	<b>4</b> out of 10	<b>↑</b> 1
Science	<b>29%</b> Meet or Exceed Standards 2016	<b>28%</b> Meet or Exceed Standards 2017	<b>-1.0%</b>	<b>7</b> out of 10	<b>→</b> 0
Reading	<b>38%</b> Growth Percentile 2016	<b>32%</b> Growth Percentile 2017	<b>-6.0%</b>	<b>4</b> out of 10	<b>→</b> 0
Graduation Rate	<b>84%</b> Graduated Class of 2016	<b>84%</b> Graduated Class of 2017	<b>0.0%</b>	<b>5</b> out of 10	<b>→</b> 0

Levels 8, 9 and 10)



Average (Index Levels 4, 5, 6, 7)



Below Average (Index Levels 1, 2, 3)

CORE DASHBOARD – CHANGES IMPLEMENTED

# District-Level Results: 2019-2020

	Metric Result 2019	Metric Result 2020	Change in Metric Performance from 2019 to 2020	Index Level 2020	Change in Index Level from 2019 to 2020
Math	<b>54%</b> Meet or Exceed Standards 2019	No Data	No Data	No Data	No Data
Language	<b>27%</b> Growth Percentile 2019	No Data	No Data	No Data	No Data
	<b>28%</b> Meet or Exceed Standards 2019	No Data	No Data	No Data	No Data
	<b>15%</b> Growth Percentile 2019	No Data	No Data	No Data	No Data
Rate	No Data	<b>89.8%</b> Graduated Class of 2020	No Data	<b>6</b> out of 10	No Data
	No Data	<b>53.8%</b> Completed A-G 2020	No Data	<b>7</b> out of 10	No Data

Levels 8, 9 and 10)



Average (Index Levels 4, 5, 6, 7)



Below Average (Index Levels 1, 2, 3)

# *How did COVID year missing data affect chronic absenteeism?*

---

- As previously discussed, after quality control and methodological review, the teams decided to report chronic absenteeism on the dashboard for 2019-2020
- Question to consider: how do we responsibly report the limitations of this data?



**Decision:** Add a tooltip to the chronic absenteeism metric (as well as other similarly impacted metrics) to clarify the business rule and data limitations for the 2019-20 school year

# District-Level Results: 2019-2020

Chronic Absenteeism in the dashboard is based on enrollment days for students between the first day of school and the last date of reported data for the 2019-20 school year (may vary by school district)

Chronic Absenteeism in the dashboard is based on enrollment days for students between the first day of school and the last date of reported data for the 2019-20 school year (may vary by school district)

Metric	Metric Result 2019	Metric Result 2020	% Change	Index Level 2019	Index Level 2020	Change in Index Level from 2019 to 2020
Chronic Absenteeism *	16.7%	16.7%	0.0%	5	5	0
Suspension Rates (includes students suspended and/or expelled) *	0.9%	0.9%	0.0%	8	8	0
Culture and Climate: FAMILY Overall	92.9%	92.9%	0.0%	7	7	No Data
Culture and Climate: STAFF Overall	82%	82%	0.0%	4	4	0
Culture and Climate: STUDENT Overall	76%	76%	0.0%	3	3	1
Social-Emotional Skills: Growth Mindset	80%	72%	-8.0%	3	3	3
Social-Emotional Skills: Self-Efficacy	71%	59%	-12.0%	5	5	4
Social-Emotional Skills: Self-Management	72%	65%	-7.0%	2	2	4
Social-Emotional Skills: Social Awareness	72%	67%	-5.0%	2	2	3



# Summary of user-centered design considerations

---

- For metrics that had unique business rules for 2019-2020, a **tooltip** (denoted with an asterisk) was included that provided the BRule definition
- Decisions were made to **completely hide the frontend view of some pages** because with missing data it would not be informative and could cause confusion
- We noted, “**No Data**” for pages with metrics that had no data
  - For 2019-20, metrics were limited to Enrollment, HS Graduation and A-G Completion, Attendance, Student Behavior, and SEL / Culture Climate

# *Partner communication & supplemental support*

- Our partners at CORE Districts developed a [2019-20 CORE Dashboard Guidance](#) document summarizing the front-end changes due to COVID-19
- We also provided supplemental analytics to the data collaborative:
  - New Rally Analytics Platform to support school re-entry and recovery
  - Produced additional analytics and reports for LEA members



## *Missing data considerations for your team to ask about front-end visualizations and reporting*

- 1 • What type of visualization is currently being displayed on the relevant dashboard?
- 2 • Is the missing data set within an historical context of other data or is it independent?
- 2 • How is the missing data interpreted between the frontend dashboard and the backend database?
- 4 • Is there a flag available within the data identifying zero or missing data or is the data just omitted?

# *Looking ahead*

# *What's next?*

---

## SY2020-21 & beyond...

- We know school closures and hybrid instruction continued through Spring 2021
- Given the assessment waivers and other COVID-related policy-decisions, it is important to continue to assess and communicate data quality issues
- District leaders need to be able to measure and understand where students are as they re-enter schools in order to implement interventions that can successfully address learning recovery and acceleration.

# *Our goal*

---

EA is committed to exploring innovative approaches to serve the data and analytics needs of our partners and their students and sharing this learning

# *Thank you!*

---

If you are interested in connecting with EA on our work in this space, please reach out to us!

Selina Eadie Dhillon ([seadie@edanalytics.org](mailto:seadie@edanalytics.org)) and

Kale Mabin ([kmabin@edanalytics.org](mailto:kmabin@edanalytics.org))



110 E. Main Street, Ste. 1000, Madison,  
WI 53703

---

608.466.4966

[info@edanalytics.org](mailto:info@edanalytics.org)

6/8/2020