Grade 3 Reading Proficiency Background Briefing

November 2013

Agenda

- The Issue
- Background
  - The Colorado Growth Model
  - Effect Size
- Approach
- Findings
- Conclusions
- Recommendations

The Issue

- In 2012, the Colorado Legislature passed the Reading to Ensure Academic Development (READ) Act
- The Act "focuses on early literacy development for all students, and especially for students at risk for not achieving third grade reading proficiency"
  - From READ Act Fact Sheet
- The READ Act focuses on Kindergarten through third grade literacy development
- The Jeffco Board has asked District Staff for an estimate of what it would cost to achieve 85% Grade 3 reading proficiency
  - Based on the 2013 TCAP, 79.6% of Jeffco third graders are at least proficient in reading (9.1% scored at the Advanced level, and 70.5% at the Proficient level)
- This Briefing is intended to help the public evaluate the District's response to the Board's request

Background

- This analysis is based on publicly available TCAP testing data, which are available on the CDE's School View website (via the Data Lab function)
- TCAP testing begins in Grade 3; results from K-3 assessments performed using reading-specific instruments (e.g., PALS, DIBELS, or DRA2) are not available to the public
- As such, the data available to the public for TCAP Grade 3 reading shows us the end-result of all the initiatives that have been undertaken over the four years from kindergarten through the spring administration of TCAP tests to third graders
- In evaluating the TCAP data, it is important that we start with a solid understanding of two concepts: (a) the Colorado Growth Model; and (b) the Effect Size metric

The Colorado Growth Model

- CSAP/TCAP scale scores measure progress along the novice to expert spectrum
- "Cut Scores" used to classify students' achievement as unsatisfactory, partially proficient, proficient, and advanced increase in each grade. Cut scores rise ever year
- "Growth Percentile" is a standardized measure that compares the increase in a student's scale score to the increase achieved by other students who all had the same scale score the previous year
- "Median Growth Percentile" (MGP) is the "Growth Percentile" achieved by the middle student in any grouping of students (e.g., a class, a school, a district)

The Same MGPs Can Reflect Different Absolute Scale Score Increases

Across Colorado, the distribution of TCAP scores for a grade should be approximately normal (i.e., bell-shaped). In normal distributions, the mean (average) score of the distribution is equal to the median (midpoint) score. Therefore, the grade-to-grade change in average TCAP score should be very close to the median gain in TCAP scale score that corresponds to the 50th Median Growth Percentile. This highlights a critical point to keep in mind about the MGP: If for different grades, the grade-to-grade changes in average TCAP scale scores are different, then two students can be in the same growth percentile, yet have very different absolute increases in TCAP scale scores. Put differently, the same Median Growth Percentile can represent very different degrees of movement up the novice to expert scale that we measure with TCAP scores. To make this more concrete, I've attached a summary of grade-to-grade changes in CSAP/TCAP Scale Scores by subject from 2006/07 to 2012/13. As you can see, the absolute score gains from Grade 8 to Grade 9 are very low compared to the score gains from Grade 6 to Grade 7. As a result, the fact that the Grade 7 and Grade 9 students were both in the 60th growth percentile does not mean that, in the absolute sense, they made equivalent progress up the novice to expert spectrum.

Grade-to-Grade Increases in TCAP Scores

Change in average state CSAP/TCAP scale score, from grade to grade								
Math	*2006 to 07*	*07 to 08*	*08 to 09*	*09 to 10*	*10 to 11*	*11 to 12*	*12 to 13*	*Average*
3 to 4	27.07	32.06	28.10	33.86	24.76	35.18	28.92	29.99
4 to 5	30.50	28.79	26.71	28.00	28.42	25.89	25.00	27.62
5 to 6	17.23	17.68	19.66	21.66	19.34	15.54	22.52	19.09
6 to 7	27.86	11.37	24.33	11.71	22.44	21.99	27.84	21.08
7 to 8	21.82	11.15	24.50	13.69	23.90	16.29	15.79	18.16
8 to 9	7.60	11.12	0.36	5.40	0.93	-0.15	2.48	3.96
9 to 10	10.36	16.05	10.51	19.36	13.73	17.52	17.09	14.95
Reading	*2006 to 07*	*07 to 08*	*08 to 09*	*09 to 10*	*10 to 11*	*11 to 12*	*12 to 13*	*Average*
3 to 4	30.21	30.51	32.49	26.54	32.86	27.48	24.80	29.27
4 to 5	21.54	29.13	24.77	26.04	25.18	25.34	24.90	25.27
5 to 6	11.34	16.91	13.76	17.94	14.00	18.68	18.92	15.93
6 to 7	12.88	14.53	10.89	13.79	11.36	14.39	11.42	12.75
7 to 8	11.46	15.73	8.83	13.63	11.42	10.59	10.27	11.70
8 to 9	9.99	14.05	7.26	14.10	6.37	7.90	8.48	9.74
9 to 10	25.48	22.17	24.14	21.71	18.23	23.22	25.28	22.89
Writing	*2006 to 07*	*07 to 08*	*08 to 09*	*09 to 10*	*10 to 11*	*11 to 12*	*12 to 13*	*Average*
3 to 4	15.01	19.00	18.90	17.08	26.03	16.70	20.52	19.04
4 to 5	22.50	27.10	20.25	22.07	25.59	17.44	23.01	22.56
5 to 6	16.12	19.12	17.78	15.70	22.09	10.42	14.44	16.52
6 to 7	31.70	23.79	32.54	24.49	29.98	29.49	37.55	29.94
7 to 8	10.26	3.33	9.85	6.47	11.92	10.92	6.17	8.42
8 to 9	7.33	3.15	6.28	1.01	3.41	1.17	8.13	4.35
9 to 10	15.09	8.49	17.87	8.90	12.54	7.54	12.21	11.80

What is Effect Size?

Effect Size (ES) is a measure of how well a system (be it a school or district) is able to continuously improve its performance over a period of time. The mathematical formula for my ES calculations is (Average TCAP Scale Score in 2013 less Average Scale Score in 2009) divided by the Standard Deviation of 2013 Scale Scores. The ES metric recognizes that improved performance can come from both higher average scores and from scores that are less variable around the average (i.e., more consistent). In his review of over 800 studies of academic performance improvement, Professor John Hattie found that the average grade-to-grade improvement in test scores was equal to about 33% of a standard deviation (see his book, Visible Learning). On this basis, he concluded that any initiative, or set of initiatives, that raised scale scores by more than .30 standard deviations was worthwhile, as it represented at least an additional year of progress along the novice to expert scale (which in Colorado we measure with TCAP scores).

More to the point for the Jeffco Board, we cannot substantially increase the percentage of students scoring at the proficient and advanced levels of TCAP or PARCC in the absence of significant positive Effect Sizes.

Relationship Between Median Growth Percentile (MGP) and Effect Size (ES)

It is also important to understand the relationship between Effect Size and Median Growth Percentile. Effect Size is based on the absolute change in scale scores over a five-year time horizon for students at a constant level (E/M/H) or grade. In contrast, Median Growth Percentile is a relative measure based on the median absolute scale score increase in a single year from grade-to-grade. It is *critical* to note that MGPs can be above the 50th percentile each year without generating a significant positive Effect Size. For example, assume that in each of five years, incoming students have a previous grade TCAP score of XXX. Further assume that in each of five years, a school raises the median student's scale score by 50 points to YYY, which results in a median growth percentile of 60 each year. Because there was no increase in YYY (the ending scale score) over the five-year period, the Effect Size would be zero. This highlights that while the school in this example is doing a commendable job with different classes of students each year, over the full five year period it has not, as a system, improved its ability to move them up the novice to expert spectrum, as there is no growth in the average ending TCAP scale scores (note that in this example I assume that the median score equals the average score, as is the case in a normal distribution).

I believe that annual Median Growth Percentiles are a useful way to measure and compare school or district value added over a one-year time horizon. However, in order to get a more accurate picture of system performance over a longer time horizon, we need to use the Effect Size metric.

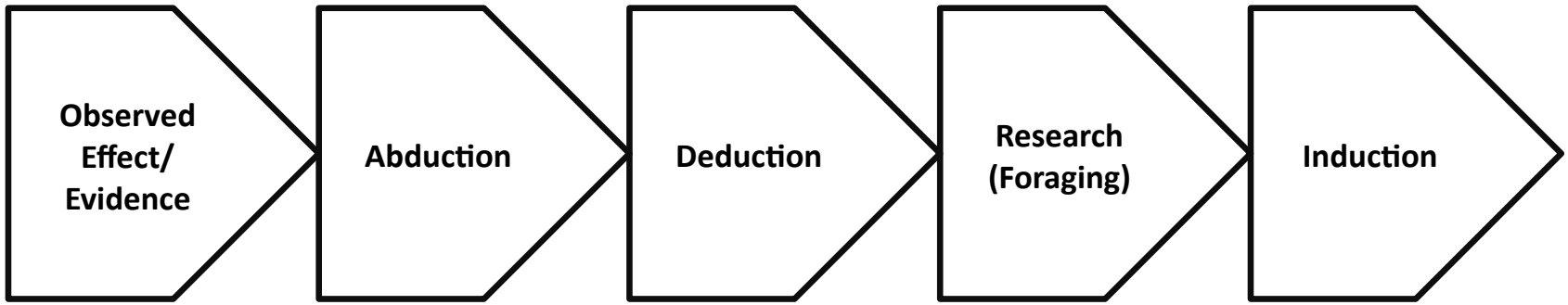
Approach

- The District and Board face a common decision problem: How to allocate scarce resources to maximize the probability of achieving a target goal (in this case, achieving 85% Grade 3 Reading Proficiency)
  - This problem is made more challenging because the metric we use to measure goal achievement will be changing (from TCAP to PARCC), and the underlying Proficiency standard will also become more challenging
- Implicitly or explicitly, boards and individuals tend to follow a four step decision process:
  - First, describe the most important features of the current situation
  - Second, explain what has caused the situation we observe
  - Third, given this causal model, predict the result of different possible courses of action
  - Fourth, on the basis of some implicit or explicit criteria, decide on the course of action to pursue
    - Note that the use of explicit criteria becomes more common when more time is available to a decision maker, and when it will be necessary to explain/justify a decision to others

Approach (cont'd)

- In complex social systems (like K12) causal explanations and predictions are usually much more difficult than they are in most physical or mechanical systems
- For example, it is hard to repeat experiments in a social system under the same conditions, because the system itself is usually constantly evolving
- As a result, in social systems, causal explanations and predictions will almost always be incomplete and subject to some unavoidable residual uncertainty about their degree of accuracy
- In social systems, causal reasoning tends to follow the process shown on the next slide

A Simplified Causal Reasoning Process



E.g. performance shortfalls, anomalous data, near misses, large or rapid change

Generate insights about possible causal explanations

If a possible explanation is true, what other evidence should we observe or not observe?

For each possible explanation, search for additional evidence with highest diagnostic value

Use collected evidence to test possible explanations and reach conclusions about their respective probabilities

Abduction Is The Most Difficult Step

- The range of causal explanations that come to mind is limited by our knowledge and experience (hence, it pays to be curious)
- It can also be constrained by our judgment about what constitutes a "plausible" (hence, it pays to be imaginative)
- Organizational limits on "acceptable" explanations can also inhibit causal reasoning (e.g., if I say this, will I offend my boss, challenge important organizational assumptions, etc.)
- For example, when explaining Grade 3 TCAP Reading results, possible explanatory factors could include:
  - Differences in curriculum used in elementary schools
  - Differences in the materials available in classrooms (e.g., books)
  - Differences in the use of technology
  - Differences in staffing (e.g., how many reading specialists at a school?)
  - Differences in teacher qualities (experience, training received, skill as evaluated by peers using SB 191 excellent teaching rubric)
  - Differences in the amount and type of support provided to students (e.g., do they have Jefferson County library cards? Has the teacher/school reached out to the parents/guardians of at-risk students, etc.)
  - Differences in the amount and type of support provided to teachers (peer evaluations, coaching, professional development received in reading skills, etc.)

Findings

- This section presents a review of Jeffco's Grade 3 TCAP Reading scores
- It is based on a number of different perspectives:
  - Differences across student groups
  - Differences between Jeffco and Cherry Creek (the District most demographically similar to Jeffco, in terms of its relative mix of different student groups)
  - Performance over time
  - Differences across schools
- The Findings will be presented as answers to different questions the public might ask
  - One set focused on where the problems are
  - And another focused on the system's rate of improvement between 2009 and 2013, as measured by Effect Sizes

On the 2013 TCAP, What Percent of Students Were Proficient or Advanced in Reading?

Grade	Jeffco P&A Percent	Of which, Advanced	Creek P&A Percent	Of which, Advanced	Comment
3	79.6%	9.1%	79.6%	10.1%	
4	76.0%	5.5%	77.1%	7.1%	
5	77.7%	11.2%	78.9%	12.7%	
6	83.1%	16.6%	70.2% **	13.2% **	** 14% of tests were not scored
7	75.9%	11.6%	74.4%	11.3%	
8	74.0%	10.6%	73.8%	11.6%	
9	72.4%	4.3%	73.4%	4.7%	
10	74.8%	12.3%	76.1%	13.9%	
11 (Average ACT Score for District)	21.3		21.7		21 is 55th national percentile

How Much Does a Student's Scale Score Have to Increase to Move From Grade 3 Unsatisfactory or Partially Proficient to the Lowest Grade 4 Proficient Score?

- For Grade 3 Reading, the midpoint scale score for Unsatisfactory is 308, and for Partially Proficient, 496 (based on the 2012 TCAP Technical Manual; these "cut scores" are consistent over time)
- The minimum Grade 4 scale score for Proficient is 572 (a gain of 76 from the Grade 3 PP midpoint, and 264 from the Unsat midpoint)
- From 2007 to 2013, the average scale score gain from Grade 3 to Grade 4 Reading was 29.97, across all students
- Therefore, a move from the Grade 3 Unsat midpoint to the Grade 4 minimum Proficient score takes about 9 times the average annual increase in scale score. From the Grade 3 PP midpoint it requires 2.6 times the average increase in scale score
- This makes it clear how important it is to get students to a Proficient Reading Standard by Grade 3, as the scale of the catch up challenge increases non-linearly with every passing grade
  - The minimum cut score for Proficient increases every grade…
  - While the average scale score gain from grade to grade tends to decrease over time, particularly between Grades 5 to 9

What Percent/Number of Grade 3 Students in Jeffco and Cherry Creek With IEPs Scored in the Unsatisfactory and Partially Proficient Categories on the 2013 TCAP Reading Test?*

Group	Jeffco/ Unsat	Jeffco/ Partial	Jeffco/ Total	Creek/ Unsat	Creek/ Partial	Creek/ Total
Female/Free & Reduced Eligible/IEP	40.0% (53)	36.0% (27)	**76.0% (80)**	41.8% (23)	38.2% (21)	**80.0% (44)**
Female/ NotF&R/IEP	21.6% (16)	37.8% (28)	**59.4% (44)**	23.5% (20)	30.6% (26)	**54.1% (46)**
Male/F&R/IEP	40.8% (53)	31.5% (41)	**72.3% (94)**	54.5% (67)	23.6% (29)	**78.1% (96)**
Male/NotF&R/ IEP	22.4% (41)	24.6% (45)	**47.0% (96)**	21.1% (34)	28.6% (46)	**49.7% (80)**

* Percentages are rounded

What Percent (Number) of Grade 3 Students Within Different Groups Scored in the Unsatisfactory and Partially Proficient Categories on the 2013 TCAP Reading Test?*

Group	Jeffco/ Unsat	Jeffco/ Partial	Jeffco/ Total	Creek/ Unsat	Creek/ Partial	Creek/ Total
F&R/Minority/ ELL	21.9% (68)	29.3% (91)	**51.2% (159)**	25.4% (86)	24.2% (82)	**49.6% (168)**
F&R/Min/ NotELL	10.2% (81)	24.5% (195)	**34.7% (276)**	11.6% (55)	23.9% (113)	**35.5% (168)**
F&R/NotMin/ ELL	18.4% (7)	15.8% (6)	**34.2% (13)**	26.6% (17)	21.9% (14)	**48.5% (31)**
F&R/NotMin/ NotELL	9.2% (82)	18.2% (162)	**27.4% (244)**	8.7% (27)	16.2% (50)	**24.9% (77)**
NotF&R/Min/ ELL	6.2% (8)	16.2% (21)	**22.4% (29)**	9.0% (17)	14.3% (27)	**23.3% (44)**
NotF&R/Min/ NotELL	4.6% (30)	14.1% (91)	**18.7% (121)**	3.7% (27)	10.3% (76)	**14.0% (103)**
NotF&R/ NotMin/ELL	0% (0)	22.0% (11)	**22.0% (11)**	3.2% (2)	14.3% (9)	**17.5% (11)**
NotF&R/ NotMin/NotELL	2.2% (71)	9.0% (283)	**11.2% (354)**	2.4% (46)	8.0% (151)	**10.4% (197)**

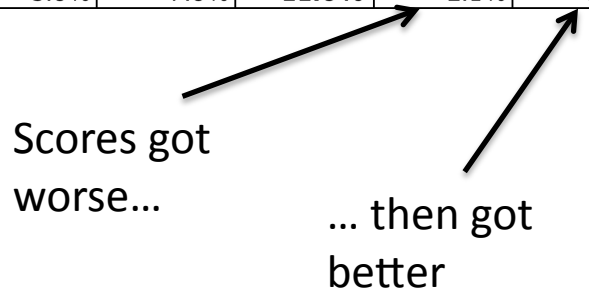
* Percentages are rounded

Which Student Groups Account for the Greatest Percentage/ Number of All Grade 3 Students in Jeffco Who Scored Unsatisfactory or Partially Proficient on the 2013 Reading TCAP?

	2013 TCAP Reading			**Number**	**Percent**
Grade 3	FARM Eligible	Minority	ELL	159	13.2%
Grade 3	FARM Eligible	Minority	NON-ELL	276	22.9%
Grade 3	FARM Eligible	Non-Minority	ELL	13	1.1%
Grade 3	FARM Eligible	Non-Minority	NON-ELL	244	20.2%
Grade 3	Not FARM Eligible	Minority	ELL	29	2.4%
Grade 3	Not FARM Eligible	Minority	NON-ELL	121	10.0%
Grade 3	Not FARM Eligible	Non-Minority	ELL	11	0.9%
Grade 3	Not FARM Eligible	Non-Minority	NON-ELL	354	29.3%
			Total	**1,207**	**100.0%**

Looking Back Three Years, Starting with Students Who Were in Grade 5 in 2013, How Have Reading Problems Evolved in Jeffco Over Time?

Jeffco	Grade 3 in 2011			Grade 4 in 2012			Grade 5 in 2013				
Group	**Unsat**	**Partial**	**U&P**	**Unsat**	**Partial**	**U&P**	**Unsat**	**Partial**	**U&P**	**Grade 3 to Grade 4 U&P Change in Jeffco**	**Grade 4 to Grade 5 U&P Change in Jeffco**
F&R/Minority /ELL	12.8%	34.1%	***46.9%***	17.7%	36.6%	***54.3%***	20.4%	27.2%	***47.6%***	7.4%	-6.7%
F&R/Min/Not ELL	15.6%	25.7%	***41.3%***	16.4%	30.2%	***46.6%***	19.9%	26.3%	***46.2%***	5.3%	-0.4%
F&R/NotMin/ ELL	17.6%	20.6%	***38.2%***	15.4%	35.9%	***51.3%***	12.5%	34.4%	***46.9%***	13.1%	-4.4%
F&R/NotMin/ NotELL	8.6%	18.4%	***27.0%***	8.3%	26.1%	***34.4%***	11.5%	19.6%	***31.1%***	7.4%	-3.3%
NotF&R/Min/ ELL	8.8%	13.2%	***22.0%***	7.1%	23.2%	***30.3%***	7.6%	17.2%	***24.8%***	8.3%	-5.5%
NotF&R/Min/ NotELL	4.5%	11.3%	***15.8%***	3.0%	15.5%	***18.5%***	4.1%	12.1%	***16.2%***	2.7%	-2.3%
NotF&R/Not Min/ELL	6.3%	10.4%	***16.7%***	4.8%	9.5%	***14.3%***	4.3%	10.6%	***14.9%***	-2.4%	0.6%
NotF&R/Not Min/NotELL	2.8%	8.5%	***11.3%***	2.6%	10.8%	***13.4%***	3.8%	7.8%	***11.6%***	2.1%	-1.8%



Looking Back Three Years, Starting with Students Who Were in Grade 5 in 2013, How Have Reading Problems Evolved in Cherry Creek Over Time?

Creek	Grade 3 in 2011			Grade 4 in 2012			Grade 5 in 2013				
Group	**Unsat**	**Partial**	**U&P**	**Unsat**	**Partial**	**U&P**	**Unsat**	**Partial**	**U&P**	**Grade 3 to Grade 4 U&P Change in Creek**	**Grade 4 to Grade 5 U&P Change in Creek**
F&R/Minority /ELL	18.1%	28.8%	***46.9%***	21.3%	36.5%	***57.8%***	20.3%	24.9%	***45.2%***	10.9%	-12.6%
F&R/Min/Not ELL	10.1%	23.8%	***33.9%***	15.3%	28.2%	***43.5%***	14.3%	22.7%	***37.0%***	9.6%	-6.5%
F&R/NotMin/ ELL	16.7%	22.9%	***39.6%***	22.9%	29.2%	***52.1%***	18.9%	17.0%	***35.9%***	12.5%	-16.2%
F&R/NotMin/ NotELL	7.8%	19.3%	***27.1%***	9.4%	25.2%	***34.6%***	10.9%	20.9%	***31.8%***	7.5%	-2.8%
NotF&R/Min/ ELL	7.6%	13.5%	***21.1%***	8.4%	21.8%	***30.2%***	3.6%	15.6%	***19.2%***	9.1%	-11.0%
NotF&R/Min/ NotELL	2.8%	10.9%	***13.7%***	3.6%	13.5%	***17.1%***	5.0%	11.1%	***16.1%***	3.4%	-1.0%
NotF&R/Not Min/ELL	3.6%	18.2%	***21.8%***	4.0%	26.0%	***30.0%***	6.3%	14.6%	***20.9%***	8.2%	-9.1%
NotF&R/Not Min/NotELL	2.9%	7.0%	***9.9%***	3.4%	10.3%	***13.7%***	2.9%	7.4%	***10.3%***	3.8%	-3.4%

Scores also got worse…

… but got better much faster

Which Jeffco Elementary Schools Had the Highest Percentage of Grade 3 Reading Problems on the 2013 TCAP?

School	Percent Unsat	Percent Partial	Total
Thompson	14%	19%	**33%**
Westgate	6%	24%	**30%**
Lasley	6%	21%	**27%**
Swanson	7%	20%	**27%**
Bear Creek	3%	20%	**23%**
Vanderhof	6%	16%	**22%**
Pennington	6%	15%	**21%**
Stevens	11%	9%	**20%**
Weber	3%	17%	**20%**
Wilmore Davis	3%	17%	**20%**
Eiber	7%	12%	**19%**
Lumberg	6%	13%	**19%**
Molholm	8%	11%	**19%**
Prospect Valley	6%	13%	**19%**
Stein	4%	15%	**19%**

These 15 Schools (15% of total elementary schools) account for 28% of Grade 3 students with reading problems (338)

Between 2009 and 2013, to What Extent (as Measured by Effect Size) Has Jeffco Improved Its Ability to Raise Reading Scores for Grade 3 Students with IEPs?

Student Group	Effect Size
Female/F&R/IEP	(.10)
Female/NotF&R/IEP	(.08)
Male/F&R/IEP	(.31)
Male/NotF&R/IEP	.09

These Effect Sizes capture cumulative impact (or lack thereof) of all the improvement initiatives that were undertaken in 2010, 2011, 2012, and 2013 with the goal of raising Grade 3 TCAP Reading scores. While we cannot tell from the public data what those initiatives were, or what their individual impact was, we can still measure their cumulative effect.

Between 2009 and 2013, to What Extent (as Measured by Effect Size) Has Jeffco Improved Its Ability to Raise Reading Scores for Grade 3 Students in Different Groups?

Student Group	Effect Size
F&R/Minority/ELL	(.12)
F&R/Min/NotELL	.02
F&R/NotMin/ELL	.05
F&R/NotMin/NotELL	(.07)
NotF&R/Min/ELL	.24
NotF&R/Min/NotELL	.02
NotF&R/NotMin/ELL	.32
NotF&R/NotMin/NotELL	.05

Remember: An Effect Size of .30 represents roughly a year of scale score gains

Between 2009 and 2013, to What Extent (as Measured by Effect Size) Has Jeffco Improved Its Ability to Raise Reading Scores for Grade 3 Students in the Most Challenged Schools?

School	Effect Size (F&R)	Effect Size (Not F&R)
Thompson	(0.37)	(0.34)
Westgate	0.26	(0.12)
Lasley	(0.10)	
Swanson	(0.19)	
Bear Creek	0.03	(0.09)
Vanderhof		(0.45)
Pennington	(0.34)	
Stevens	(0.13)	
Weber	0.06	(0.07)
Wilmore Davis	(0.20)	(0.01)
Eiber	(0.32)	
Lumberg	(0.11)	
Molholm	0.16	
Prospect Valley	(0.31)	0.03
Stein	(0.18)	

An Obvious Question is The Extent to Which the District Understands the Factors that are Driving the Differences in Effect Sizes Between the Challenged Schools and These Top Performers

School	2009-2013 Effect Size	Student Group
Belmar	0.61	F&R
Edgewater	0.35	F&R
Westgate	0.26	F&R
Deane	0.18	F&R
Foster	0.17	F&R
Green Mountain	0.16	F&R
Molholm	0.16	F&R
Stott	0.96	Not F&R
Green Mountain	0.90	Not F&R
Excel Academy	0.87	Not F&R
Mount Carbon	0.74	Not F&R
Compass Montessori WR	0.73	Not F&R
Kendrick Lakes	0.72	Not F&R
Warder	0.56	Not F&R

Causes/Explanations of Differences in Effect Sizes Could Include Some or All of These Factors:

- Differences in curriculum used in elementary schools between Grades K and 3
- Differences in the materials available in classrooms (e.g., books)
- Differences in the use of technology
- Differences in staffing (e.g., how many reading specialists at a school?)
- Differences in teacher qualities (experience, training received, skill as evaluated by peers using SB 191 excellent teaching rubric)
- Differences in the amount and type of support provided to students (e.g., do they have Jefferson County library cards? Has the teacher/school reached out to the parents/guardians of at-risk students, etc.)
- Differences in the amount and type of support provided to teachers (peer evaluations, coaching, professional development received in reading skills, etc.)

Conclusions

- Reaching 85% Grade 3 Reading Proficiency will be a challenge, as the movement from TCAP to PARCC will "raise the bar" for meeting the Proficient Standard (to a level more in line with the National Assessment of Educational Progress standard, and the international PISA standard)
  - Drops in the percent of proficient students in states like New York which have piloted the move to PARCC have been substantial
- Currently, the population of Jeffco Grade 3 students who fall short of the Proficient standard is concentrated in certain student groups and schools
- Disturbingly, the path followed by students who were in Grade 3 in 2011 shows that reading problems grew worse in Grade 4 before they were successfully addressed in Grade 5
  - Also, Cherry Creek delivered much larger reductions between Grade 4 and 5 in the percentage of students below Proficient in Reading (Why?)
- As measured by Effect Size, between 2009 and 2013, Jeffco as a system showed little progress in its ability improve Reading scores for Grade 3 students across most student groups, and in the schools with substantial percentages of Grade 3 students below the Proficient standard
- The good news is that there are examples of very substantial positive Effect Sizes in some elementary schools; whether the District has attempted to systematically understand the causal factors behind exceptionally strong and weak elementary school reading performance is an open question

Recommendations

- The Board has asked the District Staff to present estimates of what it would cost to take the action steps required to raise Grade 3 Reading Proficiency to 85% (about a 5% gain from the current level, using the TCAP Proficiency standard)
- Explicitly or implicitly, the District's recommendations will be based on predictions of the expected impact of the proposed action on the metric used to measure Reading proficiency
- There are a number of possible bases for these predictions:
  - Intuition and anecdotal experience (a questionable guide for an organization with a billion dollar budget)
  - Causal analysis of the results of successful and unsuccessful initiatives undertaken by the District, or a similar analysis of initiatives undertaken by Jeffco and other districts, like Cherry Creek
  - The results of academic analyses of the effectiveness of reading improvement initiatives
    - However, these analyses should be viewed with caution. The "Gold Standard" for this research is set by the "What Works Clearinghouse" of the Institute of Educational Sciences. For example, they recently evaluated 166 studies that had been done on the impact of the "Reading Mastery" K-6 program, and concluded that "none of these studies meet WWC evidence standards for quality research."
- It is therefore important to ensure that the District provides the Board with a clear description of the Grade 3 Reading performance problem, and its likely causes, before proposing potential solutions, and reviewing the bases for predicting their success
- In addition to ensuring that the District's recommendations are backed by clear logic, the Board should also insist that (1) the District present multiple options; (2) each option should include not only an assessment of its likely impact on Proficiency, but also its cost, timing, relative risk and potential sources of failure; and (3) a clear explanation of why it is recommending one option over the others