December 7, 2013

Mr. Ken Witt
President, Jefferson County Board of Education

Mr. Bill Bottoms
Citizen Chair
Strategic Planning and Advisory Council (in its role as District Accountability Committee)
Jefferson County Schools

Dear Mr. Witt and Mr. Bottoms,

I am the chair of the Wheat Ridge High School Accountability Committee. Pursuant to Colorado Revised Statutes, Title 22, Article 11, Section 402, we have a duty to send to you a copy of the spending priorities (and the logic that underlies them) that we have recommended to Griff Wirth, the principal of our school. I am sending you this letter in fulfillment of that duty.

To ensure that our input is as useful to you as possible, I will begin with a description of the approach we have followed in our work, summarize our key findings, and end with our conclusions and recommendations.

Approach

A list of our meeting agendas is attached. In addition to the information provided on the standard CDE school performance reports, we obtained additional data to use in our analyses, from both CDE's SchoolView/DataLab site, the Colorado TELL survey site, and CDE's ACT results site. We disaggregated this data into student groups that were mutually exclusive and collectively exhaustive (e.g., "female, eligible for free and reduced meals, not GT, and not IEP").

A major limitation on the extent of our analysis was imposed by the lack of detailed school level cost accounting data, and in particular data based on the type of activity based costing approach used by many private sector organizations.

We compared WRHS' performance on a variety of metrics in math, science, reading, and writing to 28 other high schools in the Boulder Valley, Cherry Creek, and Jeffco school districts, with special emphasis on the three that are most demographically similar to Wheat Ridge's student mix: Centaurus, Overland, and Pomona.

We focused on two metrics in particular. The first was the median growth percentile (MGP) on the most recent TCAP tests (note that these are not available for TCAP science, due to its different testing approach). However, as explained in more detail in the attached document ("Median Growth Percentile versus Effect Size"), MGP has some notable limitations as a measure of the ability of a system to improve its performance over time. For that reason, we also calculated Effect Sizes for the 2009 to 2013 period. Mathematically, Effect Size is equal to the ending average TCAP scale score for a group

of students, less the beginning scale score, divided by the ending scale score deviation. This reflects the fact that from a system performance perspective, it is not just an increase in scale score that represents an improvement in performance, but also a decrease in the variation of scale scores. Put differently the Effect Size metric expresses improvement in a school's ability to move students up the novice to expert spectrum (which we measure between Grades 3 and 10 using TCAP scores) in terms of standard deviations. Per the work of John Hattie, in K12 applications, Effect Sizes of .30 or greater are considered significant, as the average grade-to-grade increase in scale score is equal to about .30 times the starting grade's standard deviation. Hence, if the cumulative impact of all the initiatives undertaken between 2009 and 2013 increased the average scale score by more than .30, the impact of those initiatives would be equal to more than one additional year of learning.

One metric that we deliberately did not focus on was the "Catch Up" metric created by the No Child Left Behind law. In the SAC's judgment, this metric is inappropriate for a high school, as it effectively measures how well they make up for any weaknesses in the schooling received during the 9 years (K to 8) before a child enters Grade 9. Our analogy is that this metric is akin to judging an anchor leg relay runner by whether or not she can run a 30 second 400 meters in order to make up for the slow times of the three runners that preceded her.

For the same reason, we did not use as a metric the percentage of Grade 9 and Grade 10 students at Wheat Ridge who scored proficient and/or advanced on the TCAP tests, or the Grade 11 ACT test. The reason is simple: at the high school level, student proficiency is best viewed as an outcome produced by both family socio-economic status and the District as an overall system, and not solely by a high school after either 2 (in the case of TCAP) or 3 (in the case of the ACT) years of teaching. To give you a concrete example of what we mean, the high school that ranked first out of 29 for improvement in ACT scores between 2008 and 2013 wasn't Cherry Creek, or Evergreen, or Fairview. It was Jefferson High School. To be sure, too many students at Jefferson still score below the proficient level. But from our perspective, Jefferson High School has been doing a great job, even if Jeffco as a system has not.

After constructing our metrics, we moved from describing Wheat Ridge's performance to developing some hypotheses about its root causes. However, unlike the approach suggested in the CDE Unified Improvement Planning Process Handbook, we started not with performance shortfalls, but rather with those areas where Wheat Ridge's performance was exceptionally strong. Our goal was to identify performance drivers that could be matched against the results of analyses done elsewhere in the District, in the hope of identifying scalable best practices that could accelerate achievement improvement. We then moved on to root cause analysis of areas where WRHS had performance shortfalls.

Following this root cause analysis, we used our causal hypotheses as the basis for evaluating the logic and likely efficacy of the major improvement initiatives described in the school's proposed 2013-2014 Unified Improvement Plan. One unique feature of our approach was a "pre-mortem" analysis, in which we asked the WRHS leadership team

to imagine that it was a year from now, and their major improvement initiatives had all failed. What caused this to happen? What early warning signs did we miss? And what could have been done differently to avert failure? Pre-mortem analysis has been shown to be highly effective in reducing action plan failure risk in a range of military, government, and private sector contexts. We hope that it has the same impact in K-12.

Findings

Situation Description

The following table shows Wheat Ridge's mix of student categories, as well as the mixes at the three schools that are most similar to Wheat Ridge in our 29 high school comparison set:

Demographics Based on 2013 Math TCAP (GR 9&10) Test							
				Centaurus	***Pomona***	***Wheat Ridge***	***Overland***
			Students:	***1,000***	***1,480***	***1,308***	***2,200***
Female	FARM Eligible	IEP	GT	0.0%	0.0%	0.0%	0.1%
Female	FARM Eligible	IEP	Not GT	1.9%	2.1%	1.5%	1.8%
Female	FARM Eligible	Not IEP	GT	1.0%	0.7%	2.3%	0.9%
Female	FARM Eligible	Not IEP	Not GT	15.1%	17.2%	22.4%	24.8%
Female	Not FARM Eligible	IEP	GT	0.0%	0.0%	0.2%	0.0%
Female	Not FARM Eligible	IEP	Not GT	1.2%	1.1%	1.2%	1.2%
Female	Not FARM Eligible	Not IEP	GT	4.1%	4.4%	3.3%	1.8%
Female	Not FARM Eligible	Not IEP	Not GT	23.8%	23.3%	18.5%	17.1%
Male	FARM Eligible	IEP	GT	0.2%	0.0%	0.2%	0.1%
Male	FARM Eligible	IEP	Not GT	2.1%	1.6%	4.2%	4.3%
Male	FARM Eligible	Not IEP	GT	1.2%	1.3%	0.8%	1.7%
Male	FARM Eligible	Not IEP	Not GT	14.7%	17.5%	21.1%	24.9%
Male	Not FARM Eligible	IEP	GT	0.4%	0.0%	0.2%	0.3%
Male	Not FARM Eligible	IEP	Not GT	3.5%	3.3%	1.7%	2.9%
Male	Not FARM Eligible	Not IEP	GT	8.1%	5.3%	4.8%	3.2%
Male	Not FARM Eligible	Not IEP	Not GT	22.6%	22.3%	17.9%	14.9%
				100.0%	100.0%	100.0%	100.0%
	Memo Subpercentages (some double counting)						
	FARM Eligible			36%	40%	52%	59%
	IEP			9%	8%	9%	11%
	Gifted			15%	12%	12%	8%
	No Spec Category			46%	46%	36%	32%

On the 2013 TCAPs, Wheat Ridge High School delivered outstanding achievement performance, as measured by Median Growth Percentile, as can be seen in the following tables (note that these are the student groups for which publicly disclosed data is available from CDE):

Subject	F/F&R/NotGT/NotIEP	F/NotF&R/NotGT/NotIEP	F/NotF&R/GT/NotIEP
Math MGP	57	62	76*
Reading MGP	54	59	58
Writing MGP	52	59**	64*

F = female; F&R = Free & Reduced; GT = Gifted; IEP = SPED
**#1 out of 29 school comparison set; ** #2 out of 29 school comparison set*

Subject	M/F&R/NotGT/NotIEP	M/NotF&R/NotGT/NotIEP	M/NotF&R/GT/NotIEP
Math MGP	61	66	75*
Reading MGP	42	55	63*
Writing MGP	48	57	54

**#1 out of 29 school comparison set*

Compared to just our three other demographically matched high schools, Wheat Ridge's MGP ranked first in 12 out of 18 student group/subject categories.

An equally important metric is the Effect Size (ES) delivered by the Wheat Ridge team as a result of the improvement initiatives they undertook in the four years ended with the 2013 TCAP tests. We calculated 522 separate Effect Sizes (29 schools x 3 tests x 6 student groups per school). Only 58 (11%) were equal to or greater than .30. At the other end of the spectrum, 34 (7%) were less than or equal to negative (.30). These results are consistent with the general stagnation we have observed in recent years across these 29 high schools (and indeed more broadly) when it comes to increasing the percentage of students scoring at the proficient or advanced level on TCAP. The following tables show the Effect Sizes for Wheat Ridge, based on data from 2009 to 2013:

Subject	F/F&R/NotGT/NotIEP	F/NotF&R/NotGT/NotIEP	F/NotF&R/GT/NotIEP
Math MGP	.17	.24	.12
Reading MGP	.23	.11	(.17)
Writing MGP	.09	.16	(.06)

F = female; F&R = Free & Reduced; GT = Gifted; IEP = SPED

Subject	M/F&R/NotGT/NotIEP	M/NotF&R/NotGT/NotIEP	M/NotF&R/GT/NotIEP
Math MGP	.23	.25	.68**
Reading MGP	.08	.18***	.54*
Writing MGP	.11	.17***	.37**

** #1 out of 29 school comparison set*
*** #2 out of 29 school comparison set*
**** #3 out of 29 school comparison set*

We are pleased to note that 3 out of Wheat Ridge's 18 Effect Sizes (17%) were equal to or greater than .30. Compared to just our three other demographically matched high schools, Wheat Ridge's ES ranked first in 8 out of 18 student group/subject categories.

What the SAC found particularly impressive was the ability of WRHS to generate the Effect Sizes it did for students eligible for Free and Reduced Meals even as the percentage of F&R students in the WRHS population sharply increased between 2009 and 2013. Wheat Ridge's performance for boys in reading and writing was also notable, as this has been an area of particular weakness in most other high schools.

Finally, we also constructed two metrics using the 2013 ACT score data for Grade 11 students at our 29 high school comparison set, covering 2008 to 2013. Over this period, WRHS generated ACT Effect Sizes of .18 for the Composite, .19 for English, .24 for Math, .16 for Reading, and .07 for Science. We also performed a regression analysis. Just two variables – the percent of Free and Reduced and the percent of Gifted students at a high school explained (in the statistical sense, i.e., the coefficient of determination) 89 percent of the variation in the 2013 ACT Composite score for our 29 high schools. We plugged each school's values for these two variables into our regression equation to generate estimated Composite scores, and then compared them to actual scores. We express the difference in terms of standard deviations for Colorado Grade 11 ACT scores, as in the case of an Effect Size. Wheat Ridge's actual ACT Composite Score (20.5) was .16 Standard Deviations higher than predicted – the third largest outperformance in our 29 high school comparison set (behind Cherry Creek and Monarch, and tied with D'Evelyn and Fairview).

Root Cause Analysis (Explanation)

Our approach to the root cause analysis for superior performance was to start with models for high performance at professional service firms and at K-12 schools, and compared their criteria to different aspects of the Wheat Ridge system and culture committee members had observed over the past three years. We then analyzed school level data from the TELL Survey to identify areas where the answers provided by Wheat Ridge teachers were significantly different than the average answers for teachers at all Colorado high schools. On the basis of these inputs, we drew our inferences about factors contributing to Wheat Ridge's outstanding academic achievement performance.

One of the best descriptions of the characteristics of a high performance team was provided more than twenty five years ago, by US Army General Walter Ulmer: "What is the essence of a 'good climate' that promotes esprit and gives birth to 'high performing units'? It is probably easier to feel or sense than to describe. It doesn't take long for most experienced people to take its measure. There is a pervasive sense of mission. There is a common agreement on what are the top priorities. There are clear standards. Competence is prized and appreciated. There is a willingness to share information. There is a sense of fair play. There is joy in teamwork. There are quick and convenient ways to attack nonsense and fix aberrations in the system. There is a sure sense of rationality and trust. The key to this climate is leadership in general, and senior leadership in particular."

In K-12, Jeffco's Strategic Compensation pilot project has noted some of the characteristics of high performing teams in the school context. These include:

- Working in teams, to shift teachers' focus from "my kids" to "our kids"
- Analyzing data and developing shared strategies to improve student learning
- Teachers receiving more frequent, useful and specific feedback to improve instruction
- Teachers reflecting on their professional strengths and areas to improve
- Teachers getting job embedded professional development based on their individual needs
- Principals building cultures of trust and collaboration
- Principals and staff working together to develop a shared approach to leadership in their building

Our analysis of the specific responses in the TELL survey found that areas of significant difference between WRHS and the average Colorado High School were correlated with high performance team indicators from the benchmark models we used.

Wheat Ridge scored significantly higher on:

- *"Teachers have sufficient non-instructional time"*
- *"Teachers are protected from duties that interfere with their essential role of educating students"*
- *"Teachers have sufficient time for collaboration"*
- *"School leadership facilitates the use of data to improve student learning"*
- *"Time spent [by teachers] utilizing results of assessments"*
- *"Teachers are assigned classes that maximize their likelihood of success with students"*
- *"Teachers are trusted to make sound professional decisions"*
- *"Teachers have autonomy to make decisions about instructional delivery"*
- *"Teachers are encouraged to reflect on their own practice"*
- *"Teachers determine the content of professional development"*
- *"PD offerings are data-driven"*
- *"Professional learning opportunities are aligned with the school's improvement plan"*
- *"The faculty are recognized for accomplishments"*
- *"Teachers have an appropriate level of influence on decision making at this school"*
- *"Teachers are effective leaders in this school"*

Beyond this survey data, we have three further observations to make about likely root causes of Wheat Ridge's superior performance. While they are based on anecdotes (the plural of which is not data), they are grounded in committee members' years of work on performance improvement projects in private sector companies.

One of the first thing a student or parent encounters at Wheat Ridge when they arrive in Grade 9 is regular reminders about the Farmers' long "Tradition of Excellence", in the arts, athletics, and academics. This is not just a slogan; the halls are filled with evidence

of the many students who have excelled over the years at Wheat Ridge. Quite simply, students quickly get the message that at Wheat Ridge, expectations for everyone's performance are high.

The second thing both students and parents quickly notice at Wheat Ridge is that the expectation of excellence is matched by an extraordinarily high level of, and comprehensive approach to student support (at least in comparison to other schools committee members' children have attended). At one level, this springs from deliberate policy decisions, such as mandatory attendance at after school teacher access periods for students who are struggling in a class, or the explicit emphasis at Wheat Ridge on developing non-cognitive skills that are critical to both academic and life success, such as persistence, resilience, conscientiousness and grit. Yet at another level, the extent of student support one observes, whether on the individual level, or in the context of student participation in clubs, the arts, and athletics, is clearly an emergent property of the system and culture, rather than the result of a deliberate policy decision. Anyone who has had a student at Wheat Ridge will not be surprised by the TELL survey finding that the school scores much higher than the average Colorado High School on the amount of time teachers spend on both "student interactions" and on "communication with parents, guardians, and the community." In fact, one can argue that this high level of student support and external communication has produced a virtuous circle, generating both higher levels of community and parent engagement and support than are found in many schools, and teachers who rate (on the TELL survey) "community support and involvement" one of the two most important reasons they are willing to keep teaching at Wheat Ridge (the other being the high quality of teacher leadership).

Finally, it is quite obvious that everything I have written about thus far has not come about in the absence of exceptional building leadership. As already noted, strong teacher leaders are part of this story. But an equally important part is the leadership provided by the principal (Griff Wirth) and his assistant principals. There is an important subtlety here that it is critical not to overlook. Observing many aspects of life at Wheat Ridge for almost four years, I have once again been reminded that there is no single leadership style that is best for all situations; while there are some enduring qualities found in all effective leaders (e.g., integrity, competence, and empathy), the right leadership style to use always depends on the nature of the challenges facing an organization, and the mix and quality of people you have on your team. In some circumstances, a highly directive, controlling, and charismatic leadership style may be right; yet in other situations, such as the one I've seen at Wheat Ridge, a servant leadership ("upside down pyramid") approach may be optimal. Of course, this also raises the equally critical point that indiscriminate changes made to the Wheat Ridge staffing model by the District could very easily trigger a sharp reduction in organizational performance, as such changes have so often done in the private sector, and I would guess, in K-12 as well.

Prediction/Action Planning

I will now turn to the achievement growth challenges facing Wheat Ridge, their root causes, and the steps that are being undertaken to meet them.

As previously noted, over the past five years, Wheat Ridge has experienced a very significant evolution in its school demographics, with sharp growth in the percentage of students eligible for Free and Reduced Meals. This the principal root cause that underlies the challenges that are the focus of two of the three Major Improvement Initiatives found in both last year's and this year's Unified Improvement Plan.

The first of these is improving literacy. A key challenge here is that, in general, high school teachers have not expected to be required to teach basic literacy in their classes. Demographic changes are now requiring this, which in turn requires not only increased school wide collaboration and focus on this issue, but also increased instructional support. Wheat Ridge has already cut a life and consumer science teacher from its budget in order to fund a Reading Instructional Specialist. However, our Committee concluded that demand for these services is sufficiently high as to merit the addition of another Reading Specialist, and/or funding for additional experimentation with potentially more efficient blended learning/technology enabled reading instruction (with the caveat that reading and writing instruction, even with technology support, seems likely to always be relatively labor/instructor hours intensive). Unfortunately, it is not clear to us (given the absence of activity based cost data) whether it would be possible to cut a further existing staff position to make room for another Reading Specialist, especially given the strong positive impact on cognitive and non-cognitive skill growth produced by the school's excellent arts programs, and the growth in the number of IEP and GT students at the school (we also note that we have concluded that the presence of both high numbers of GT and FARM students may be more symbiotic than anyone had expected, particularly in a school with a strong culture and strong student supports, such are found at Wheat Ridge). Apart from these staffing and technology funding questions, the specific sub-initiatives planned under the Improving Literacy Major Improvement Initiative seem on target to us.

The second critical initiative that will continue into the next school year is a focus on improving the academic culture at Wheat Ridge and increasing the effectiveness of interventions based on assessment data. With respect to the latter, while the TELL survey, TCAP, and ACT growth and Effect Size results indicate that WRHS is better than most schools in its use of data to drive performance improvement, it is also the case, as it is elsewhere, that not every member of the professional staff is at the same place in this process. As such, there is an ongoing need for coaching and support in this area. Perhaps more important is the challenge of maintaining the "Tradition of Excellence" in the face of significant demographic changes. The major negative for Wheat Ridge in the TELL survey was a high level of frustration with the disciplinary situation at the school. A number of steps have been taken to address this challenge, including an explicit focus on improving the transition of middle school students into high school (and into the Wheat Ridge culture in particular), mandatory additional instruction time (Access) to support struggling students, and twice yearly goal setting conferences with students that explicitly incorporate assessment metrics and data.

While these initiatives all strike our Committee as sound, we also recognize, and would like to highlight for the DAC and the Board, that while Wheat Ridge has gone to

extraordinary lengths to adapt to a significant increase in F&R students (including many unpublicized voluntary actions by Wheat Ridge staff that as extraordinary as they are heartwarming, and which speak volumes about the school's culture), we believe that more efficient support from the District Central Office would make their lives much easier, and likely lead to faster achievement improvement. To be sure, shortfalls in the efficiency and effectiveness of head office support for field units is not unique to Jeffco, nor to K-12; in fact, it has been a recurring theme in the private sector, and elsewhere in the public sector, for years. Yet that does not mean we should simply accept it as a given. For example, we have no doubt that the lives of Wheat Ridge staff and students would be greatly improved if Jeffco could institute a single point of contact at head office for F&R support issues, which has the skill, knowledge, and authority to bring all of the system's capabilities to bear in a well coordinated manner, without forcing staff members at WRHS to make multiple calls to try to do this themselves. This would also undoubtedly generate benefits for Jeffco schools far beyond Wheat Ridge.

The third Major Improvement Initiative at Wheat Ridge is "to improve systemic use of best instructional practices through cross-content learning teams." Whether the root cause of the need for this initiative is a lack of skill or a lack of will is unclear. I also note that this initiative is fully consistent with the District's focus on this area (e.g., the introduction of the "Instructional Rounds" process). We believe that this focus has the potential to generate significant improvements in achievement. We also believe that, assuming the root cause giving rise to the need for this initiative is a lack of skill, then the sub-initiatives that Wheat Ridge is undertaking in this area, including expanded use of peer observation and coaching, and targeted professional development activities, are logically consistent and should, if implemented with high fidelity, lead to the desired improvements in this area.

Pre-Mortem Results

As previously noted, we conducted a "pre-mortem" analysis of Wheat Ridge's UIP, asking school staff to imagine that it was a year from now, that all the Major Improvement Initiatives had failed, and to tell us why that had happened (and what we could have done to avoid that outcome). The three major risk factors identified were:

- A radical change in the district's employee compensation policy, that would cause staff to "take their eye off the ball", and/or cause the loss of key staff.

- Loss of key staff (whether through resignation or transfer), and their replacement by people who did not fit well into the Wheat Ridge culture and leadership style.

- A further significant increase in the percentage of Free and Reduced eligible students at the school before the adaptations to previous increases have been fully implemented. Put differently, in every system there is a "tipping point" or "critical threshold" that triggers a phase change; Wheat Ridge's challenge is to avoid exceeding this.

Conclusions and Recommendations

- The Wheat Ridge team has delivered an outstanding performance over the past five years, resulting in substantial improvements in academic results even as they adapted to a dramatic increase in the percentage of Free and Reduced Eligible students at their school.

- The root causes of this superior performance are easy to understand in light of various models for high performance professional teams. However, while it is easy to define such models, actually implementing them is far more difficult. Yet this has been done at Wheat Ridge.

- Two of the Major Improvement Initiatives in Wheat Ridge's Unified Improvement Plan reflect the need for further adaptations to the substantial changes in the school's student demographics. With respect to improving literacy, there appears to be a strong case for providing additional funding to the school, for both increased staffing and to allow an already high performing school to experiment with potential technology and blended learning driven interventions that could further increase its rate of academic achievement growth. With respect to maintaining and further strengthening the school's academic culture, Wheat Ridge would greatly benefit from the creation of a single point of contact at the District Central Office for F&R support issues, which has the skill, knowledge, and authority to bring all of the system's capabilities to bear in a well coordinated manner, without forcing staff members at WRHS to make multiple calls to try to do this themselves. At the school level, this has no budget implications. That said, at the policy level, the District, DAC, and Board of Education would all undoubtedly benefit from a detailed, activity-based study of the true costs of adequately supporting F&R students at different levels of our K-12 system.

- We have great hope for the achievement improvement impact of Wheat Ridge's and the District's increased focus on using a range of collaborative, team-based initiatives to improve instructional practice. Unfortunately, in the absence of activity-based cost information, we were unable to discern the budgetary implications of this initiative for Wheat Ridge.

If you have any questions about this letter, please don't hesitate to contact me.

Very truly yours,

Tom Coyne

Wheat Ridge High School
Proposed SAC Meetings and Agendas for 2013/2014

Meeting	*Agenda Items*
September	• Review previous year's Unified Improvement and Performance Plans in light of most recent TCAP and other performance data • Assess success of key performance improvement initiatives from last year's plan • Prioritize performance challenges for coming year • Identify other schools which have had success in meeting these challenges, and external research that bears on them • Review ideas for increasing parent engagement in the school, and for soliciting parents to serve on the SAC, pursuant to new language added to 22-11-402 by SB-193
October	• Review root cause analyses for top priority challenges • Review initiatives that will be undertaken (and included in the UIP) to address these challenges
November	• Review Next-to-Final Draft of the UIP • Review budget and district policy implications of the UIP • Complete a pre-mortem/risk analysis of the plan as a committee • Agree on any changes to make to plan before submission
December **Performance Plan due to District on Dec 15th**	• Review final draft of Unified Improvement Plan. • Agree on report to send to District Accountability Committee and School Board (per 22-11-402(a): "The SAC for a district public school shall send a copy of its recommended spending priorities to the school district accountability committee and to the local school board." • Discuss how to effectively follow up/advocate with district and/or state with respect to policy and/or budget changes that are needed to execute UIP

Meeting	*Agenda Items*
	• Review situation with respect to parent engagement, and initiatives to improve it
January	• No meeting
February	• Meeting to discuss status of key initiatives, and how to remove any obstacles that may be negatively affecting the speed or fidelity of implementation
March	• No meeting, unless there are significant obstacles to implementation of key initiatives
April	• Meeting to discuss status of key initiatives

How to Get More Insight From Colorado Growth Model (TCAP) Data

Median Growth Percentile versus Grade-to-Grade Scale Score Increase versus Effect Size

By Tom Coyne

October 2013

The Colorado Growth Model

- CSAP/TCAP "scale scores" measure progress along the novice to expert/ learning curve spectrum
  - The TCAP scoring scale goes from 150 to 999 for Reading, and 150 to 950 for Writing and Math
  - In theory, a student starts as a novice in Grade 3 (the first TCAP test grade), and progresses up the learning curve from there to Grade 10, the last TCAP test year
- "Cut Scores" are used to classify students' achievement as unsatisfactory, partially proficient, proficient, and advanced
  - Cut scores rise every year
  - E.g., for Grade 3 Reading, the minimum score for proficient is 526; by Grade 10, the minimum score for proficient is 663
- "Growth Percentile" is a standardized measure that uses at 100 point scale to compare the increase in a student's scale score to the increase achieved by other students who all started with 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 for State of Colorado

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

Why Grade-to-Grade Scale Score Increase is Important

- Each year, as a student moves from grade to grade, the minimum TCAP "cut score" that he or she must achieve to be deemed "Proficient" or "Advanced" increases
- If the average grade-to-grade increase in scale score is less than the increase in cut score, a student can have an impressive growth percentile (which is based on the former), but still fail to move up an achievement category (which is based on the latter)
- This is the answer to a frequently asked question: "If we are delivering such impressive median growth percentiles, why aren't we also seeing an increase in the percentage of students who are proficient or advanced?"

In the following slides, I'll use Jeffco as an example of this.

Here are Jeffco's Median Growth Percentiles for the Past Eight Years

Jeffco TCAP Data										
Math	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Median Growth Percentile
3 to 4	47	45	49	45	49	49	50	54		49
4 to 5	50	48	55	54	56	52	54	52		53
5 to 6	58	56	58	61	61	61	61	61		60
6 to 7	50	50	53	56	51	55	58	55		54
7 to 8	49	51	58	55	53	52	50	51		52
8 to 9	47	49	54	55	53	57	54	53		53
9 to 10	51	50	57	55	51	56	54	57		54
Reading	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Median Growth Percentile
3 to 4	51	56	51	51	53	54	51	52		52
4 to 5	46	47	49	49	49	52	49	52		49
5 to 6	54	58	60	60	60	63	60	60		59
6 to 7	45	48	47	48	45	48	47	47		47
7 to 8	44	48	53	47	48	48	49	49		48
8 to 9	44	49	51	49	50	52	46	45		48
9 to 10	42	50	50	51	48	54	46	50		49
Writing	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Median Growth Percentile
3 to 4	52	51	52	51	50	54	50	53		52
4 to 5	48	45	48	46	45	48	46	47		47
5 to 6	56	51	55	52	57	60	58	58		56
6 to 7	47	47	47	45	45	44	45	47		46
7 to 8	45	48	45	44	45	46	48	47		46
8 to 9	46	49	49	48	49	48	48	47		48
9 to 10	48	52	54	52	50	50	50	49		51

If you look just at Median Growth Percentiles, Jeffco students are keeping up with the rest of the state (and doing better in math).

Here are Jeffco's Gains in Average Scale Score, Compared to the Increase in the Cut Scores for Proficient

Jeffco TCAP Data												
Math	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Student Increase	Incr in Min Proficient Scale Score	Jeffco Avg/Min Prof Incr
3 to 4	23.59	17.84	14.61	21.12	30.89	23.37	34.09	30.66		24.52	36.00	68%
4 to 5	36.29	28.90	33.89	30.76	33.27	30.64	29.19	25.31		31.03	39.00	80%
5 to 6	18.43	22.17	18.54	27.48	28.22	28.12	22.69	29.35		24.38	26.00	94%
6 to 7	9.20	24.37	16.72	25.26	8.20	22.61	22.96	28.16		19.69	39.00	50%
7 to 8	10.79	19.95	24.49	25.55	12.71	22.96	13.32	12.86		17.83	18.00	99%
8 to 9	13.25	7.90	15.93	2.93	7.38	4.60	2.58	5.92		7.56	25.00	30%
9 to 10	14.14	7.97	17.83	10.48	14.17	14.48	16.80	19.76		14.45	25.00	58%
Total Gain	*125.69*	*129.10*	*142.01*	*143.58*	*134.83*	*146.79*	*141.64*	*152.03*		*139.46*	*208.00*	
Reading	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Student Increase	Incr in Min Proficient Scale Score	Jeffco Avg/Min Prof Incr
3 to 4	31.87	34.81	27.96	29.63	24.05	30.36	23.80	22.85		28.17	46.00	61%
4 to 5	22.52	19.96	25.65	23.90	25.01	26.14	24.84	26.21		24.28	16.00	152%
5 to 6	13.26	17.64	23.81	20.58	24.24	22.24	23.55	23.25		21.07	12.00	176%
6 to 7	6.63	10.59	7.97	6.61	6.67	6.34	7.25	5.51		7.20	20.00	36%
7 to 8	9.21	10.07	16.61	7.30	11.51	10.82	8.70	9.22		10.43	12.00	87%
8 to 9	6.85	9.69	13.67	4.12	13.90	6.12	4.32	5.39		8.01	10.00	80%
9 to 10	19.45	24.57	21.78	22.52	19.80	19.74	19.37	26.65		21.73	21.00	103%
Total Gain	*109.80*	*127.32*	*137.47*	*114.66*	*125.19*	*121.77*	*111.84*	*119.08*		*120.89*	*137.00*	
Writing	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Student Increase	Incr in Min Proficient Scale Score	Jeffco Avg/Min Prof Incr
3 to 4	13.74	12.59	18.83	17.65	15.26	28.32	14.45	21.66		17.81	20.00	89%
4 to 5	22.20	18.04	26.36	16.44	18.09	24.19	14.25	21.70		20.16	13.00	155%
5 to 6	23.49	16.97	25.61	20.40	22.67	31.96	17.40	21.04		22.44	15.00	150%
6 to 7	16.92	28.99	21.08	26.61	21.28	22.96	23.17	33.50		24.31	26.00	94%
7 to 8	4.38	10.43	-0.10	5.86	5.83	10.70	10.56	5.07		6.59	17.00	39%
8 to 9	8.79	9.26	2.89	7.70	5.53	4.47	2.52	8.87		6.25	7.00	89%
9 to 10	8.65	15.52	11.99	16.93	8.57	10.52	6.15	11.10		11.18	15.00	75%
Total Gain	*98.17*	*111.80*	*106.67*	*111.60*	*97.22*	*133.12*	*88.50*	*122.93*		*108.75*	*113.00*	

In contrast to Median Growth Percentiles, scale score gain data show that Jeffco students are often falling short of the achievement growth required to stay above the proficient cut score.

Here are the Percentages of Jeffco Students Scoring Proficient and Advanced Over the Past Eight Years

Jeffco TCAP Data										
Math	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Pct Proficient & Advanced
Grade 4	72	73	71	72	75	75	77	77		74
Grade 5	70	67	70	67	70	71	68	70		69
Grade 6	65	68	67	71	69	71	71	70		69
Grade 7	53	58	55	61	56	61	61	63		59
Grade 8	54	55	56	61	58	59	58	59		58
Grade 9	47	44	48	45	50	47	43	46		46
Grade 10	38	39	40	40	39	42	42	43		40
Reading	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Pct Proficient & Advanced
Grade 4	74	72	72	72	75	73	75	76		74
Grade 5	75	73	77	74	75	78	77	78		76
Grade 6	77	79	80	81	82	82	84	83		81
Grade 7	72	72	71	73	75	74	75	76		73
Grade 8	73	70	75	69	74	73	74	74		73
Grade 9	72	73	73	75	73	72	72	72		73
Grade 10	70	74	72	75	70	71	71	75		72
Writing	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Pct Proficient & Advanced
Grade 4	60	57	60	57	58	63	56	58		59
Grade 5	68	62	65	62	62	66	63	63		64
Grade 6	69	68	68	69	66	71	65	67		68
Grade 7	64	68	65	67	65	64	67	68		66
Grade 8	57	59	58	57	60	59	59	60		59
Grade 9	58	57	56	59	56	58	56	59		57
Grade 10	54	57	55	56	53	53	52	53		54

Grade-to-Grade Scale Score Gain clearly has a tighter linkage to proficiency than Median Growth Percentile.

Here are the Proficient and Advanced Percentages Just for Students Who are Not Eligible for Free and Reduced Meals

Jeffco TCAP Data										
Math	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Pct Proficient & Advanced
Grade 4	79	80	79	80	83	83	86	85		82
Grade 5	77	74	78	76	79	80	78	81		78
Grade 6	72	75	74	79	78	79	80	79		77
Grade 7	60	64	63	70	66	72	73	74		68
Grade 8	61	62	63	69	67	69	69	71		66
Grade 9	53	50	54	52	57	56	52	57		54
Grade 10	43	44	45	46	45	49	50	52		47
Reading	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Pct Proficient & Advanced
Grade 4	80	79	81	80	83	81	85	85		82
Grade 5	83	80	84	82	84	86	85	87		84
Grade 6	84	85	86	88	88	89	91	90		88
Grade 7	78	79	78	80	83	84	84	85		81
Grade 8	79	76	81	77	82	82	84	83		80
Grade 9	78	78	79	81	80	80	80	82		80
Grade 10	75	79	77	81	77	78	79	82		79
Writing	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Pct Proficient & Advanced
Grade 4	67	65	69	67	68	72	67	69		68
Grade 5	76	70	74	72	72	75	74	74		73
Grade 6	76	75	75	78	76	80	74	77		76
Grade 7	71	74	72	76	74	74	77	78		74
Grade 8	64	66	65	66	69	69	70	71		67
Grade 9	65	62	63	66	63	68	66	70		65
Grade 10	59	63	60	62	61	60	61	63		61

Eliminating F&R Students from the data doesn't make the performance problem go away. And it was there before the 2009 budget cuts.

You Also See the Same Performance Problems for Free and Reduced Students – Only they are Worse

Jeffco TCAP Data										
Math	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Pct Proficient & Advanced
Grade 4	53	53	51	53	57	59	60	61		56
Grade 5	48	43	47	45	49	53	49	50		48
Grade 6	43	47	47	51	49	54	53	51		49
Grade 7	28	33	27	35	32	38	39	42		34
Grade 8	29	28	31	34	34	36	35	36		33
Grade 9	22	21	23	20	27	23	21	24		23
Grade 10	16	15	19	16	17	21	20	20		18
Reading	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Pct Proficient & Advanced
Grade 4	55	48	51	51	58	55	57	59		54
Grade 5	53	50	56	55	56	60	60	60		56
Grade 6	56	60	63	62	67	67	71	70		64
Grade 7	50	48	49	51	54	54	57	58		53
Grade 8	49	46	52	45	52	50	54	56		51
Grade 9	47	51	51	51	53	50	53	53		51
Grade 10	46	50	49	53	47	52	51	57		51
Writing	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013		Average Pct Proficient & Advanced
Grade 4	37	33	35	35	35	44	34	37		36
Grade 5	44	37	39	39	39	46	44	41		41
Grade 6	47	46	46	47	46	53	46	50		48
Grade 7	40	45	39	42	41	42	47	49		43
Grade 8	33	33	33	32	35	35	37	38		34
Grade 9	30	31	31	32	32	33	32	37		32
Grade 10	29	30	30	31	27	30	29	32		30

In sum, when trying to improve achievement, focusing on Median Growth Percentiles distracts you from critical dynamics and performance issues.

What is Effect Size?

Effect Size 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.

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 30% 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 to undertake, as it represented at least an additional year of learning progress up the novice to expert scale (which in Colorado we measure with TCAP scores).

The mathematical formula for my five year ES calculations is (Average TCAP Scale Score in 2013 less Average Scale Score in 2008) 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).

My ES metric is based on public data about TCAP outcomes, and not the result of any specific achievement improvement initiative. As such, it captures the cumulative impact of all the achievement improvement initiatives that were undertaken over a given period of time on the TCAP or ACT scores for students at a given level of our educational system.

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.

5 Year Effect Sizes for Jeffco, Based on 2008 and 2013 TCAP Results, are Weak; None > .30

5 Yr Effect Sizes (08 to 13) for All Students	Math	Reading	Writing
Grade 3	(0.02)	0.05	(0.07)
Grade 4	0.14	0.04	(0.06)
Grade 5	0.01	0.01	(0.12)
Grade 6	0.10	0.08	(0.04)
Grade 7	0.27	0.10	0.13
Grade 8	0.10	(0.03)	0.04
Grade 9	(0.01)	(0.09)	0.08
Grade 10	0.10	0.07	0.00

The clear implication of these Effect Sizes is that, at the aggregate District level (if not at individual schools), the impact of all the achievement improvement initiatives undertaken over the past five years has been minimal. As a result, the 2013 TCAP results showed that substantial percentages of Jeffco tenth graders were not proficient in math, writing, and reading. And PARCC will significantly raise the bar for proficiency.

	2013 Grade 10 TCAP Results for Jeffco	
	Free and Reduced	*Not Free and Reduced*
NOT Proficent in Math	**80%**	**48%**
NOT Proficient in Writing	**68%**	**37%**
NOT Proficient in Readin	**43%**	**18%**

What is the Relationship Between TCAP 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.

5 Year Effect Sizes for Jeffco, Based on 2008 and 2013 Grade 11 ACT Test Results. None > .30

	Jeffco Grade 11 ACT 5 Year Effect Sizes				
	Composite	English	Math	Reading	Science
2008	20.4	19.8	20.3	20.5	20.6
2013	21.2	20.6	21.1	21.3	21.1
Std Dev 2013*	5.4	6.6	5.2	6.5	5.4
Effect Size	0.15	0.12	0.15	0.12	0.09
	* standard deviation for Colorado				

Note that the Grade 11 ACT Test is the last standardized assessment that every student in Colorado must take. It is therefore the final comprehensive and comparable metric we have to evaluate the impact of District initiatives on achievement improvement