An Evidence Based Approach to Improve College and Career Readiness

By Tom Coyne

How can we move beyond the ideological differences that characterize too many discussions about education today, and get on with the critical work of substantially improving Colorado's student achievement results?

Just as evidence-based medicine has led to faster improvement in healthcare outcomes, so too can an evidence-based approach to education produce better results for our children.

Let's start with how to define the achievement improvement goals we want our school districts to meet. The last comprehensive measure we have of the cumulative result of our investment in K-12 education is the ACT assessment that is taken by every 11th grader in Colorado (which next year will be replaced by the SAT). Not only is the ACT important for college admissions, its results are also highly correlated with other tests that students may take, including the Armed Services Vocational Aptitude Battery (ASVAB), and the WorkKeys assessment for students seeking a National Career Readiness Certificate.

The ACT establishes benchmark scores for "college and career readiness" in different subject areas. Consider the 2015 results for students not eligible for the free and reduced lunch program who live in six relatively affluent suburban districts: Boulder Valley, Cherry Creek, Douglas County, Jefferson County, Littleton, and St. Vrain Valley. Only 52% of these students met the ACT's "college and career ready" standard in reading, only 54% in math, and only 50% in science.

The results for students eligible for free and reduced lunch were much worse.

Let's assume our goal is to have 75% of all 11th grade students meet or exceed the ACT's college and career ready benchmarks in reading, math, and science. How much would a district's average score have to improve to reach that target?

If you assume that a district's ACT scores are normally distributed (i.e., they look like the familiar "bell curve" when plotted on a graph), then it is a relatively straightforward calculation.

Things get much more difficult when you ask people to divide scarce budget dollars across different initiatives to close this achievement gap. How are these decisions typically made in your district?

Too often, they seem to result from a combination of ideology and political power, with an occasional smattering of outside research evidence. Unfortunately, this approach has not produced substantial gains in student achievement performance despite the billions of taxpayer dollars we spend each year on our public schools.

Is there a better way to approach this problem? As a starting point, we need a common metric that both describes the size of the achievement gap we are trying to close, as well as the relative effectiveness of different achievement improvement initiatives we could pursue. This metric exists but it is too seldom used, much less used systematically to drive continuous improvement over multiple years.

Rather than specifying the size of the achievement gap in terms of absolute score points on the ACT or SAT, it is much more useful to use a standardized measure, in order to make it comparable not only with achievement gaps based on other tests, but also with research findings about the effectiveness of different achievement improvement initiatives.

One way to do this is to divide the size of the achievement gap expressed in score points (i.e., the average ACT score if 75% of our students met the college and career ready benchmarks, less the actual 2015 score) by the standard deviation of the 2015 scores (standard deviation is a measure of how widely those scores are distributed around their average). The resulting metric -- the size of the gap expressed as a multiple of the standard deviation -- is also known as the "effect size."

The following table converts 2015 district achievement gaps in reading, math, and science in twelve Front Range districts into effect sizes:

Required Improvement in Average ACT Scores for 75% of Students to Meet College & Career Ready Benchmarks On Grade 11 ACT, Expressed in Standard Deviations (Based on 2015 ACT Results)

District	Reading	Math	Science
Academy	0.57	0.71	0.77
Boulder Valley	0.46	0.49	0.60
Cherry Creek	0.74	0.67	0.93
Cheyenne Mountain	0.38	0.28	0.58
Colorado Springs	1.13	1.24	1.35
Douglas County	0.71	0.73	0.82
Falcon	1.13	1.39	1.37
Jefferson County	0.80	0.88	0.97
Lewis Palmer	0.33	0.49	0.58
Littleton	0.55	0.58	0.69
Poudre	0.55	0.67	0.75
St. Vrain Valley	0.88	1.01	1.04
Average	0.68	0.76	0.87
ACT Std Dev. (State)	*6.40*	*5.30*	*5.50*

Because district level standard deviations aren't publicly available, we've used the state level data. To convert an effect size back to ACT scale score points, multiply it times the standard deviation at the bottom of the column. For example, the average size of the ACT reading gap equals .68 x 6.40, or 4.35 points.

Now that we've established the size of the achievement gap we want to close (over some period of time), the next step is to look at research findings about the effect sizes for different achievement improvement initiatives that a district could pursue. Our goal is to identify a mix of achievement improvement initiatives that maximizes the expected effect size for whatever budget we have available, recognizing that the full realization of these effect size gains will take a number of years.

At this point, we confront another problem: the lack of replication of many of the findings reported by education researchers (e.g., see, "*Facts are More Important than Novelty: Replication in the Education Sciences*" by Makel and Plucker). Actually, this problem is much broader, as replicability of research findings in a growing number of disciplines is increasingly being called into question (e.g., see, "*Research Reproducibility, Replicability, Reliability*" by Ralph Cicerone, president of the United States National Academy of Sciences).

The solution to this problem is to not rely on the effect sizes reported in individual research studies, but rather on those reported from so-called "meta-analyses", which combine the results of multiple studies on the basis of the strength of the methodologies they use. Fortunately, when it comes to student achievement improvement initiatives, there is no shortage of meta-analyses we can use.

Let's start by examining the effect-sizes for four achievement improvement initiatives that are frequently encountered in district Unified Improvement Plans: Smaller Classes, Early Childhood Education, Response to Intervention (a methodology for systematically delivering more instructional support to struggling learners) and more investment in Teacher Professional Development.

The Washington State Institute for Public Policy (WSIPP) is the non-partisan arm of the Washington State Legislature that is charged with conducting meta-analyses and cost-benefit assessments of various policy proposals. Their analyses found an effect size (ES) of just .01 for reducing kindergarten class size, and .007 for reducing first grade class size. However, these were broad conclusions; higher effect sizes were found for class size reductions targeted at younger, at-risk, and male students. Their general conclusion

was as follows:

"First, the weight of the evidence indicates that, on average, class size is related to student outcomes—smaller class sizes improve outcomes, although the overall effect appears to be small. Second, the positive effect of lowering class size is much stronger in lower school grades and weaker in the upper grades. The bottom-line finding from our analysis of the evidence and economics of class size reduction is that in the earliest K–12 grades reducing class size has a high probability of producing a favorable outcome—that is, where the long-term benefits of reducing class size consistently exceed the costs. In the upper grades, on the other hand, reducing class size poses a substantial risk of an unfavorable outcome—that is, where costs may often exceed benefits" ("*K-12 Class Size Reductions and Student Outcomes*").

It is also important to keep in mind – and this is a key point in the larger district context – that the impact of an intervention on the overall district achievement gap equals its expected effect size times the size of the treated student population relative to the overall district. For example, an effect size of .20 for a group of students who account for 30% of a district would generate a .06 reduction in the overall district achievement gap.

What about increased investment in Early Childhood Education (ECE)? In their meta-analysis ("*Early Childhood Education for Low Income Students*"), WSIPP found an effect size on achievement outcomes (in grades K – 2) of . 152, that faded to .085 by grades 6 – 9. Another WSIPP meta-analysis of full day kindergarten for disadvantaged students found an initial effect size of .12, which "faded out to nearly zero by grades two through five" ("*Full Day Kindergarten: A Review of the Evidence and Benefit-Cost Analysis*").

Response to Intervention (RTI) is, at first glance, a very common sense approach that uses assessment results to target additional instruction ("Tier 2" and "Tier 3") at students who most need this support. Initial research on RTI pilots found very impressive effect sizes. For example, in "*Meta-Analytic Review of Responsiveness-to-Intervention Research: Examining Field-Based and Research Implemented Models*", Burns et al found a median effect size of 1.09 for this approach (see also, the Institute for Educational Sciences' *Practice Guides for RTI* in Reading and Math). Given this initial promise, over

70% of U.S. school districts now use some form of RTI.

Unfortunately, scaled up implementation of RTI has sometimes failed to live up to the results achieved in the early pilots. For example, in November 2015 the U.S. Department of Education's Institute for Education Sciences published an updated meta-analysis of RTI reading interventions ("*Evaluation of Response to Intervention Practices for Elementary School Reading*") that found zero to negative effect sizes for the 20%-30% of students who received additional Tier 2 support.

What about more investment in teacher professional development (PD)? In "*Teacher Compensation and Training Policies*", WSIPP's meta-analysis found an effect size of zero for general professional development programs, and . 005 for content specific PD. These findings are consistent with other studies that have found that teacher professional development investments have at best a very minimal impact on student achievement results. For example, The New Teacher Project recently found that the districts they studied spent an average of $18,000 per year per teacher on professional development with no positive return (see their report, "The Mirage").

These PD findings are frustrating, because other research has found substantial effect sizes for the student achievement gains that result from having a highly effective teacher instead of one who is just average. In "*Measuring the Impacts of Teachers*", Chetty et al found that in New York, the effect size in elementary school was .12 for English Language Arts (ELA) and .16 for math, while in middle school it was .08 for ELA and .13 for math.

In a study of the impact of highly effective teachers in Los Angeles ("*Validating Teacher Effect Estimates Using Changes in Teacher Assignments in Los Angeles*"), Bacher-Hicks et al found effect sizes of .19 and .29 for ELA and math in elementary school, and .10 and .21 for ELA and math in middle school. Increasing the percentage of highly effective teachers in our schools is clearly an important leverage point for improving student achievement.

Unfortunately, researchers have yet to identify all the factors that drive superior teacher effectiveness. For example, in "Teacher Compensation and Training Policies", WSIPP's meta-analyses identified the effect size impacts of the following teacher factors on student achievement results:

- Having a master's degree = -.004
- Having a graduate degree in the subject taught = .023
- Individual Pay for Performance = .005
- Intensive Induction Programs = .07

In sum, when you compare the size of the achievement gaps districts need to close with the predicted effects of the most common initiatives they are pursuing, it is clear that they are insufficient to meet the challenge we face.

Are better approaches for closing student achievement gaps available? Let's take a look at the effect sizes for a non-exhaustive list of other initiatives that a district could pilot.

One of the most powerful steps that a district can take is to change the curriculum it uses in a given subject. For example, the Institute of Educational Sciences' meta-analysis of different math curricula found that switching from Investigations to Math Expressions had an effect size of .30 ("*Achievement Effects of Four Early Elementary School Math Curricula*"). Similar effect sizes have been found for different reading programs (see, "*Reviewing Systematic Reviews: Meta-Analysis of What Works Clearinghouse Computer-Assisted Reading Interventions*" by Streke and Chan).

Instructional initiatives also seem to hold promise for closing achievement gaps. For example, in "*A Meta-Analysis of Interventions for Struggling Readers in Grades 4-12*", Scammacca et al find an average effect size of .21. And in "*A Meta-Analysis of the Effects of Instructional Interventions on Students' Mathematics Achievement*" Jacobse and Harskamp find an average effect size of .58. Finally, in "*A Nation Deceived*", Colangelo et al's meta-analysis found that grade acceleration of gifted students produced a .80 effect size.

Another approach to student achievement improvement is broader adoption of the initiatives that have proven successful in public charter schools (via either their incorporation into district-run schools, or the expansion of charter schools). For example, in "*No Excuses Charter Schools: A Meta-Analysis of Experimental Evidence on Student Achievement*", Cheng et all find ELA effect sizes of .07 for ELA in elementary school, .07 in middle

school, and .21 in high school. For math, the corresponding effect sizes are . 12, .16, and .27.

In "*Injecting Successful Charter School Strategies Into Traditional Public Schools*", Fryer reports that, "We implemented five strategies gleaned from practices in achievement-increasing charter schools – increased instructional time, a more rigorous approach to building human capital of teachers and administrators, high-dosage tutoring, frequent use of data to inform instruction, and a culture of high expectations – in twenty of the lowest performing schools in Houston, Texas. We show that the average impact of these changes on student achievement is 0.206 standard deviations in math and 0.043 standard deviations in reading, per year, which is comparable to reported impacts of attending high-performing charter schools."

While the research into the effectiveness of various competence-based and personalized learning approaches that make greater use of technology is still developing, there are some early indications that this area holds great promise for student achievement improvement. For example, a meta-analysis by Kurt Van Lehn found an average effect size of .79 for adaptive computer based tutoring systems ("*The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems*").

Increased focus on developing students' social and emotional skills is another promising student achievement improvement initiative. For example, in "*The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions*", Durlak et al found an average effect size of .32 for the impact of these programs on academic performance. Similarly, in "*The Role of Noncognitive Factors in Shaping School Performance*", Farrington et al found an average effect size of .27.

Last but not least, other research has found that improving district management and governance processes can also have a strong impact on student achievement. For example, in "*School District Leadership That Works*", Waters and Marzano's meta-anlaysis found an average effect size of .24, with key underlying drivers that included:

- Establishing non-negotiable goals for achievement and instruction;

- Board alignment with and support of these goals, including allocation of sufficient resources to initiatives to meet them; and
- Regular board and district leadership monitoring of progress toward achieving the goals that had been established.

To be sure, these meta-analysis findings are just a start, and it will take time to pilot, test, and scale initiatives that are not yet being pursued by a district, as well as to understand how they interact with each other. Moreover, these initiatives also differ not only in their expected effect sizes, but also in their cost. As WSIPP has repeatedly noted, a critical challenge for school districts is allocating scarce K-12 funds to maximize the expected improvement in student achievement results over a given period of time.

It is also not enough to simply agree to pilot initiatives that research suggests could have a large positive impact on student achievement performance. Districts also need to implement these experiments effectively, rigorously and transparently evaluate their results, and commit additional funds to scale up those that meet or exceed their predicted effect sizes.

In sum, it is clear that an evidence-based approach to student achievement improvement can help more of our children graduate from high school college and career ready, while also enabling the rest of us to find common ground and move beyond the increasingly polarized debate that characterizes too many discussions about our public schools.

By taking an evidence-based approach to education, all Coloradans can win.

Tom Coyne is a member of Jeffco's District Accountability Committee, co-founded www.k12accountability.org, and has worked on corporate performance improvement issues for more than 30 years.