2016 October 16

ESSA INDICATORS

Indicators of School Quality and Student Success

CHIEFS for CHANGE

CHIEFS FOR CHANGE IS A COALITION OF STATE EDUCATION CHIEFS AND DISTRICT SUPERINTENDENTS DEDICATED TO EXCELLENCE AND EQUITY FOR ALL STUDENTS.

We advocate for the policies and practices working for students, facilitate a robust system of peer-to-peer advising among our members, and sustain a pipeline of the next generation of Chiefs.

To learn more about Chiefs for Change, visit our website at chiefsforchange.org

CHIEFS for CHANGE

CONTENTS

EXECUTIVE SUMMARY 4

KEY SECTIONS

Introduction 5

Considerations for Selecting ISQ Indicators 7

Considerations for Evaluating the Research Evidence 8

Research Overview of Non-Test-Score Indicators 13

- Student Attendance 21
- Teacher Attendance 22
- Student Suspensions 23
- School Climate 23
- Student Retention 24
- Student Engagement 25
- Non-Cognitive Skills 25
- Early Warning Indicators 26
- Credit Accumulation 27
- High School GPA 28
- International Baccalaureate Diploma (and Pre-Diploma) Program 28
- Advanced Placement Courses 29
- Dual Enrollment 30

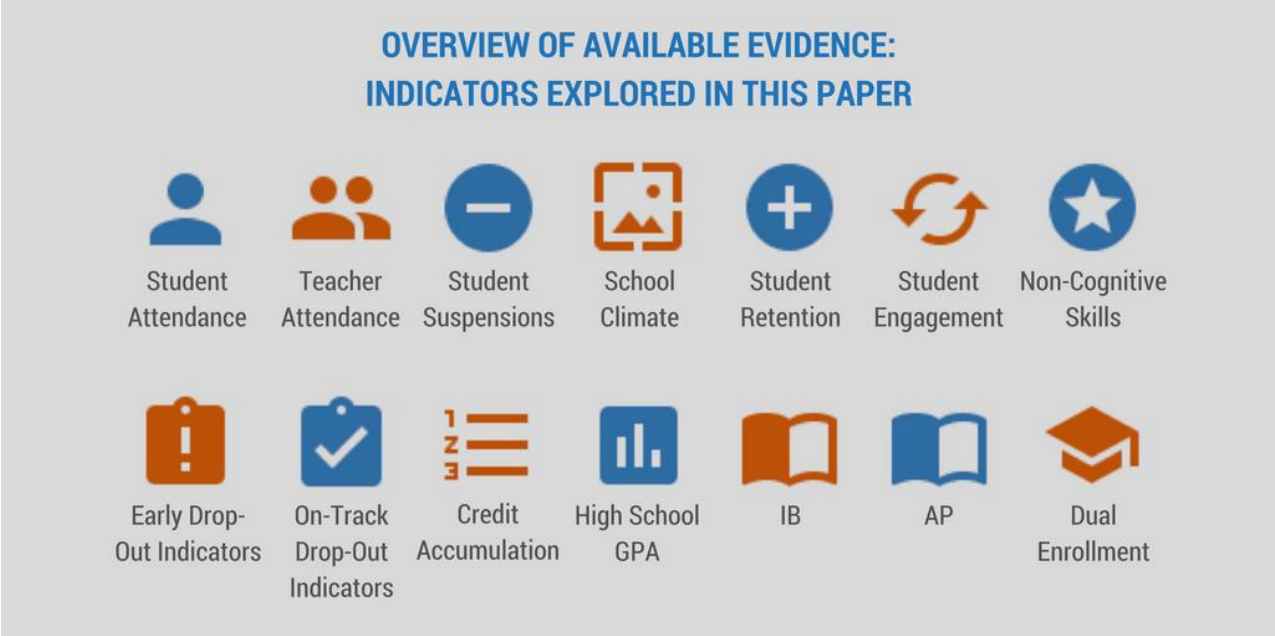
EXECUTIVE SUMMARY

State accountability frameworks provide an important opportunity to provide consistent and transparent reporting of key school information to students, parents, educators, and other stakeholders. Under ESSA, states have significant flexibility in determining the indicators they will use to evaluate schools and determining appropriate supports and interventions. One key requirement is for states to include an "indicator of school quality or student success" in their systems. While this indicator cannot be given significant weight, it does play an important role in systems to signal what the state considers important in school performance.

Many states already include one or more indicators in their systems that match this definition and may not need to make revisions based on this requirement. However, some states will consider adding or modifying indicators in this category under new ESSA systems. The purpose of this paper is to analyze the research available with regard to these potential measures so that states can consider this as a key factor in determining which measures to select. There are numerous other important factors to consider including stakeholder feedback, cost of implementation, ability to disaggregate by student population, and the extent to which the measure effectively differentiates schools.

It is important to note that some measures may not meet these criteria for inclusion in the formal school accountability system, but may still be worthwhile to report publicly to parents and educators to provide additional data and information about schools. These measures may evolve to be formal accountability measures over time, or may remain important data points to share but not include in accountability systems.

In all ESSA-related decision-making, we encourage states to use evidence as a key guidepost for choosing the best options along with other essential criteria. We hope this paper provides a brief but helpful overview of the available evidence with regard to inclusion of many potential measures that could be used for this ESSA indicator. Finally, we note that research on these and other measures continues to evolve and we encourage states to adopt processes of continuous improvement that allow them to re-evaluate and refine their systems on an ongoing basis as additional research and information becomes available.



Introduction

The Every Student Succeeds Act (ESSA) requires states to use at least one "indicator of school quality or student success" (hereafter ISQ) that "allows for meaningful differentiation in school performance" and "is valid, reliable, comparable, and statewide," alongside the other required annual assessment data in their accountability systems[1]. Schools and school districts must also be able to disaggregate data related to that indicator to inform stakeholders about students in different subpopulations: those from all racial and ethnic groups, students with disabilities, children from low-income families, and English learners.

While the ESSA requirement for an ISQ has received significant attention, it is important to note that some states already include such an ISQ in their accountability systems (e.g., drop-out rate, student attendance, credit accumulation, AP/IB participation/performance, dual enrollment, etc.). Therefore, while ESSA presents a clear opportunity for states that are interested in phasing in new indicators now or in the future, some states do not *have to* choose a brand new ISQ in order to comply with the new law, although they will almost certainly have to take new steps to ensure that they meet the sub-group reporting requirements that govern all ISQs.

Before considering new indicators, states will want to critically examine the quality and validity of all the existing indicators in their systems, particularly with regard to any performance targets in their annual growth indicators and so as to ensure that these are consistent with their ESSA-required state-wide interim and long-term academic achievement targets. Nevertheless, ESSA clearly also requires at least one ISQ as part of a state's accountability system, thus providing an important opportunity to focus on new indicators that could contribute to continuous improvement in educational outcomes.

As they consider the type and number of these new indicators to add, states should be conscious of previous history. Under waiver and prior federal accountability requirements, some states found that they had tried to measure too much at the state level, had not considered the concerns of parents and families, and had not sufficiently recognized the technical, organizational, and fiscal capacity constraints of state education agencies (SEAs) and local education agencies (LEAs).

For states that are considering new or additional indicators, this brief provides an overview of frequently discussed possibilities and a summary of the research (or lack thereof) to support their inclusion.[2] We urge states to carefully consider which indicators may be ready for inclusion in accountability systems in the near term, versus those that might be phased in over time as additional information becomes available.[3]

---

[1] Every Student Succeeds Act, Pub. L. No. 114-95, § 1111, 129 Stat. 1836, (2015).

[2] Note that the purpose of this paper is to provide research evidence on indictors of state-determined outcomes. The specific interventions needed to improve an outcome are beyond the scope of this paper.

3 Chiefs for Change is grateful to David Steiner and Alanna Bjorklund-Young at the Johns Hopkins Institute for Education Policy for providing the research and analysis cited in this paper.

There may also be data elements that a state wishes to report on, but not to include in formal school accountability determinations. States will want to distinguish between indicators that belong in a state accountability system and those which a district or school may want to collect for diagnostic purposes and share with parents and the wider public. As they move forward, states will naturally want to be sensitive to their stakeholders and responsive to local interests. In particular, both the indicators themselves and the associated results need to be expressed in language that is accessible and meaningful as they are shared with parents.

Finally, states will need to consider their capacity to meet ESSA's data disaggregation requirements. ESSA requires that all accountability indicators are valid and reliable and broken down by subpopulations, but very little research to date has investigated the most discussed indicators in this light. The process of disaggregating may be relatively straightforward for some indicators - such as student suspension rates - but for others such as school climate, data systems may need to be upgraded to meet this requirement. For example, while it is entirely feasible to report school climate through multiple surveys for subpopulations in a school, a minimum sample size to generate meaningful data needs to be determined, as well as the validity of comparing differing views from subpopulations versus the entire school.[4]

Considerations for Selecting ISQ Indicators

[4] For example, research finds that race/ethnicity is an important predictor for explaining student perceptions of school climate (Thepa et al., 2013). Thus research must determine if this is because students of different races are treated differently, have different expectations, or a combination of the two, before we understand how to best compare school climate scores from subpopulations to the entire school population.

There are three main issues to consider when selecting indicators aside from the student outcome of interest.

1. **The first critical consideration is to determine which indicators are suitable to include in accountability systems**. Some indicators are valid and reliable for research purposes, but can be sensitive to potential manipulation when attached to incentives. For example, studies show that "conscientiousness" (sometimes labeled "grit") and "mindset" can be reliably and validly measured using student surveys; however, top researchers caution against using them within accountability systems, since adults can easily influence student responses (Duckworth & Yaeger, 2015). The problem becomes still clearer in the case of suspension rates, which are highly correlated to dropout rates. But if the suspension rate became an accountability indicator, schools could be incentivized to reduce suspensions to zero without taking other steps to improve learning conditions for at-risk students, thus risking negative impacts on the academic performance of other students. Therefore, placing such measures in an accountability system, where adults have incentives and the ability to change student responses or produce simplistic solutions to complex educational challenges, should be considered with great caution.

2. **The second consideration is to choose an indicator that measures school-based factors, as opposed to factors that are largely beyond the schools' control**. Student absence is a case in point: While strong research suggests that absences correlate with academic outcomes, absences are also highly correlated with students' socio-economic backgrounds (Hansen, 2016). Therefore, a better indicator might combine an absolute measure of student absenteeism with a gap-closing measure, so that schools with chronically low attendance could show progress. Otherwise, we risk simply equating factors that are largely beyond the schools' control, i.e., an out-of-school effect, with factors that the school can control.[5]

3. Finally, while states must consider the strength of the evidentiary basis for selecting a particular indicator for their accountability system, **they may have policy preferences** which push them to choose an indicator with a relatively modest research base. For example, states might feel confident that their standardized tests accurately measure student achievement, but for political reasons desire indicators that are predictive of dropping out of high school, even if the research linking specific indicators to that outcome is relatively modest.

Considerations for Evaluating the Research Evidence

[5] States use academic growth measures alongside proficiency measures for this same reason.

The briefing memo that follows provides specific research findings connecting indicators (such as teacher absenteeism) with important student outcomes (such as college readiness). However, for reasons we will explain below, the research simply does not enable us to offer a hierarchical ranked recommendation of the discussed indicators.

We define strength of research in several ways.

ESSA requires that the ISQ indicator must at least be "valid, reliable, comparable, and state-wide" (§ 1111, 129 Stat. 1836) and the draft regulations additionally require that these indicators are "supported by research that performance or progress on such measures is likely to increase student achievement or, for measures within indicators at the high school level, graduation rates."[6] Our own recommendation is that in selecting one or more ISQs, states not only meet this requirement but determine the level of research that supports their potential choice(s). This way they will maximize the chances that the indicators will serve as successful incentives to create states' desired educational outcomes.

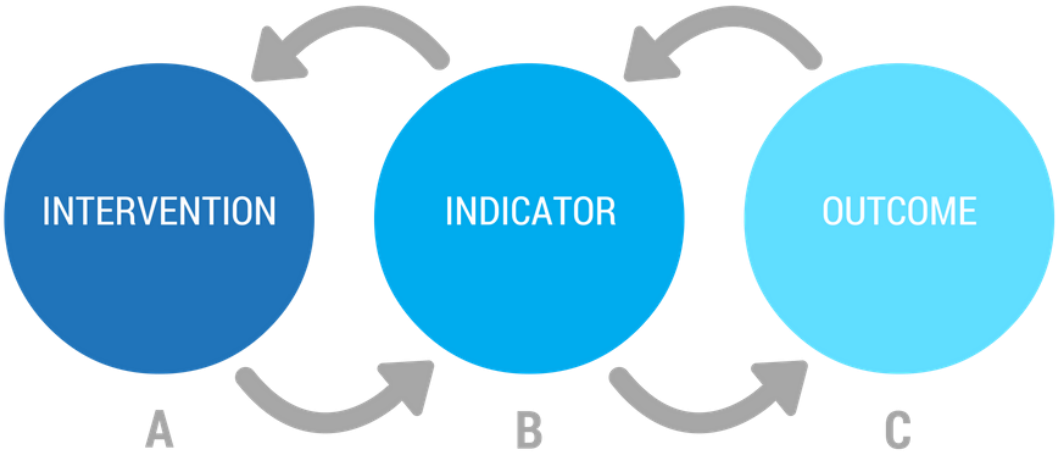
A very useful measure of the strength of research is the tiers of research evidence defined elsewhere in ESSA.

ESSA defines four levels of research evidence in the following way (§ 8002, 129 Stat. 2091):

1. Strong Evidence: based on at least 1 well-designed and well-implemented experimental study (often called a Randomized Control Trial).
2. Moderate Evidence: based on at least 1 well-designed and well-implemented quasi-experimental study.
3. Promising Evidence: based on at least 1 well-designed and well-implemented correlational study with statistical controls for selection bias.
4. Strong Theory: includes ongoing efforts to examine the effects of such activity, strategy, or intervention.

[6] The final regulations issued by USDOE governing the accountability provisions in ESSA use the language of "student learning" rather than "student achievement" to clarify that the department does not intend "achievement" to be understood as referring only to student test performance. In the remainder of this paper, we thus adopt the new language.

When considering the functional difference between the levels of research on ISQs, states should keep in mind that they are in fact dealing with three inter-related components: (A) the intervention (concrete action(s) that teachers or schools actually perform) (B) the indicator the state is considering (the data point(s) it is actually collecting for accountability purposes) and (C) the student outcome (the results the state hopes to achieve). States will likely first choose a student outcome and then establish a set of indicators that are informative about that outcome. The level of research on the indicator impacts what the intervention should target in order to achieve the desired student outcomes.



The distinction between these three components and their research base is crucial. Level 1 research is the strongest category of evidence because it indicates a causal relationship between the indicator and the student outcome (i.e., a change in the indicator *causes* a change in the student outcome). Level 1 evidence is also particularly useful for policy purposes, because randomized control trials provide evidence about the efficacy of a specific intervention that changes the indicator, thereby providing information about what a school should do to impact the indicator and thus the student outcome. *In short, only level 1 evidence directly ties all three components together.*

Unfortunately, we have almost no level 1 research for ISQ indicators. One reason for this is that few of the indicators can be studied using a randomized control trial. For example, it is not feasible to randomly select a group of students and force them to be absent for a specific amount of time, in order to study the effects of these absences on their academic achievement.

We therefore mainly find level 2 and level 3 evidence for the indicators below. To take the same example of using student attendance as our indicator, Level 2 evidence plausibly provides causal estimates of an indicator's effect on an outcome (e.g., the effect of student attendance on standardized test scores), although one has less confidence of causality than if the data had come from a randomized control trial. When data from a randomized control trial is unavailable, researchers can sometimes use statistical techniques that allow causal interpretation of the estimates.[7] Level 2 research is also less useful than level 1 for policy purposes, because this research does not always include a specific intervention that impacts the indicator. *Therefore, while level 2 research tells us that impacting an indicator will impact the associated outcome,* ***that same research*** *does not necessarily tell us how to impact that indicator – or in this case, it leaves us in the dark as how to increase student attendance. Schools and districts will have to review all the available research on how to move that indicator (or the outcome directly), and then choose the strongest research-backed intervention they can afford.*

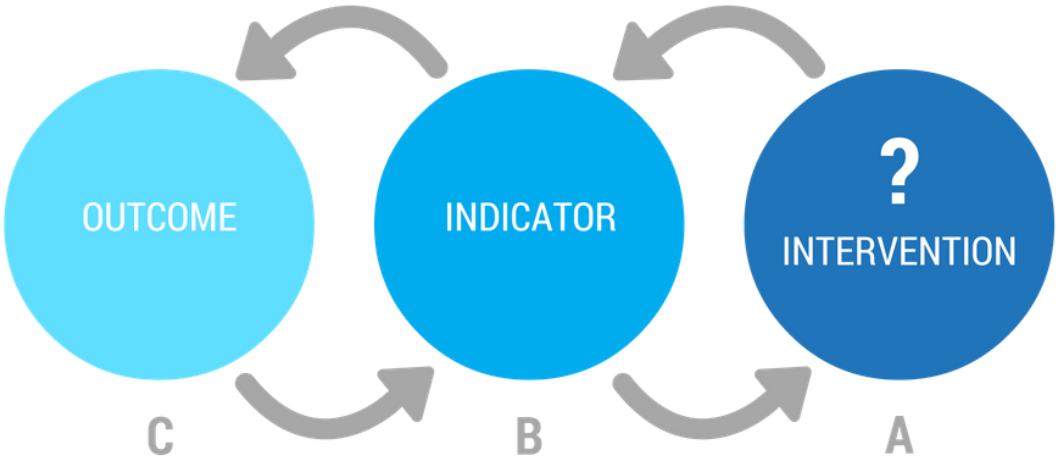
In contrast to level 2, level 3 evidence is correlational. This means that the indicator and the student outcome both change at the same time, but we cannot say that the indicator causes the student outcome to change. For example, we know that student suspensions (an indicator) are highly correlated with students' dropping out of high school (a student outcome). However, does the act of being suspended cause students to drop out? Does a specific student behavior drive this relationship? Are students acting out and becoming suspended because they are academically behind (and is it this latter fact that is the true cause of their later dropping out)? Correlational research evidence simply does not identify the underlying causes of the relationship, but rather

[7] For example, in order to study the effect of student absences on student achievement, researchers used snowfall as an instrument for school closings and student absences (Goodman, 2014). Instrumental variables are variable that are only associated with the outcome of interest (e.g. student achievement) through an observed variable (e.g. school closure), but not with variables that are unobserved to the statistician (such as student background, teacher quality, or school quality). Researchers use instrumental variables to studying the effect of student absences on student achievement in order to obtain a causal estimate, because student absences are correlated with student characteristics (such as race), some of which are plausibly not in the data.

offers only predictive information about the relationship between an indicator and the student outcome. *The risk with level 3 indicators is that schools and districts will try to find interventions that directly impact the indicator, just as they would (rightly) do for a level 2 indicator; but now that would be a mistake, since by definition, they could have no assurance that impacting the level 3 indicator will change the student outcome.* ***What they would have to do instead is look for interventions that other research suggests could impact the student outcome directly.***

ESSA's level 4 evidence is not considered statistically meaningful. We do not include any examples below.

To further clarify the distinctions between the different levels of evidence, consider the following concrete example. The state has chosen (B) – an indicator of student engagement and (C) the student outcome of high school persistence (i.e., a reduction of dropout rate). A school district is left having to find (A) – namely what actions it will support in its schools in an effort to reduce the dropout rate (C).



If the state has level 1 or 2 research about their chosen indicator (student engagement), then the district will know that increasing student engagement will increase the student outcome of high school persistence. Level 1 research will even tell the state specifically what researchers found to be effective in increasing the indicator of student engagement and thus the district can rationally employ these same actions. Neither the state nor the district would need to look for any additional research. Level 2 research establishes the causal relationship between the indicator of student engagement and the student outcome of high school persistence, but does not provide information about how to move the indicator. Therefore, the state (ideally) or the district would need to gather additional research about potentially effective interventions. However, because of the causal relationship in level 2 research, the intervention can target either the indicator or could directly focus on the student outcome.

If, on the other hand, student engagement has only level 3 research, this means that the indicator of student engagement and the student outcome of high school persistence typically move together, but does not

establish that increasing student engagement will change high school persistence. The state should therefore offer districts examples of interventions that directly target the student outcome, high school persistence. Note that there are fewer policy options with level 3 research: the state no longer has the option of targeting the indicator, student engagement. In the case of indicators supported only by level 3 research, if the state lacks any confidence that it can find strong examples of interventions that directly move the outcome, we would strongly urge the state not to adopt that indicator.

How should state policymakers think about all of this? An indicator supported by ESSA's level 2 research – showing a causal relationship between an intervention and a student outcome – is normally preferable to an indicator resting on ESSA's level 3 research – showing only a correlational relationship. This is because you want to know that if you change an indicator, an effect on the related student outcome will follow. We list below some of ESSA's level 2-ranked research that causally links teacher absenteeism to academic performance. Measuring and then reducing teacher absenteeism, if it has reached the critical level to effect student performance, could be prudent policy. School and district leaders would then search for the best research-based interventions that reduce teacher absenteeism. If we had merely correlational data linking teacher attendance with student learning, then one could not conclude that impacting the former would create a change in the latter.

One might thus conclude that state policymakers should as a rule choose an indicator with level 2 rather than level 3 research support, even if the research for the former suggests a relatively modest causal relationship and the research for the latter suggests a very strong correlational relationship. This is especially important if the indicator is to be used for formal accountability, since as we have suggested, in the absence of causation-based research about the indicator, a district and/or school would have to target the outcome directly, and doing so might not impact its "rating" on the indicator.

There are perhaps rare cases when a state might nevertheless choose a level 3 research-supported indicator. A state might be interested in using the very high predictive power of an indicator's relationship to an important outcome. For example, a state that is concerned with high school dropout rates might implement an indicator for ninth graders who are behind in credit accumulation or have failed more than one class. The correlational research suggests that such students are four times more likely to drop out of high school than their peers. Attention to these students is urgently required, and states might wish to push schools and districts to use a research-based, targeted intervention strategy for those students. *But the state would also need to make very clear to districts and schools that their intervention strategies must target the outcome itself (dropout prevention), and not the indicator.* For example, a randomized control trial provides evidence that "career academies" have a statistically significant effect on the desired outcome (high school completion rates) (Kemple and Snipes, 2000). Districts might do well to choose that intervention. But again, we would urge states interested in doing something like this not to put the level 3 indicator into their formal accountability system, since the research has not established that focusing on that indicator will cause a better measure on the outcome.

There may, of course, be non-research-based issues, including real-world politics and cost, that influence the choice of measures. Certainly, cost might make some choices of research-based ISQs unavailable, but spending less money on policies with no strong research base is unlikely to benefit students.

For these reasons, it would not be realistic to offer a cast-iron, rank-ordered list of the measures themselves. ***In general, we would recommend that, as they choose an indicator, states strongly bias their indicators towards those with strong level 2 research support whenever possible to avoid encouraging schools and districts to directly target level 3 indicators, which do not have an established causal relationship with student outcomes.*** They should also be sensitive to the grade-sensitivity of research and perhaps create a portfolio approach with interventions that vary according to the school level (elementary, middle, high). States could, for instance, include teacher attendance for elementary schools; student attendance for middle schools; and percentage taking and passing demanding course options for high schools.

Research Overview of Non-Test-Score Indicators

There are numerous indicators that states might use as a non-test-score indicator in accountability systems. The chart below provides a summary overview of the most-discussed measures, the grade levels in which they are best used, the magnitude of the effect size, and the strength of the research base to support its inclusion. Additional information on each of these indicators and supporting research is included after the chart.

Indicator	Outcome Measure	Grade Level	Exemplary Finding Summary, Effect Sizes, etc.	ESSA Level of Research Evidence (See text above for details)	Strength of Causal or Correlational Relationship[8] (1 is strong-est, 4 is weakest)
Student Attendance	Academic achievement, high school persistence	K-12	Promising: evidence suggests that each additional absence leads to a .25% to 5% standard deviation (SD)[9] decrease in students' math scores in grades 3-5.[10]	2	2

[8] Definition of "Research Evidence"

1. Strong: the results indicate substantial effects (0.25 SD or more) OR evidence of strong positive correlation (over 0.5) with outcome, strong predictive power (identified over 50% of population of interest), or positively correlated with multiple outcomes of interest. Such indicators are recommended, especially as one of multiple measures.
2. Promising: the results indicate modest effects (less than 0.25 SD), correlated with outcome (below 0.50) with outcome and less predictive power (identifies 50% of population or interest or less). Such indicators are cautiously recommended, preferably as one of multiple measures.
3. Inconclusive: mixed or inconclusive results. Not recommended.
4. Evidence needed: lack of evidence of a relationship with outcome. Not recommended.

Note that as a rule, we do not consider research evidence from the publisher of a tool or any other interested parties.
The definition of effect sizes comes from:
Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. Child Development Perspectives, 2(3), 172-177.
Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Vol. 2. Lawrence Earlbaum Associates, Hillsdale, NJ.

[9] Effect sizes provide a standardized measure of how a treatment impacts an outcome of interest. In this case, a student who misses an additional day of learning decreases their math test score by .25%-5% of a standard deviation (SD). This is in comparison to students who did not miss an additional day of school and therefore did not change their math test scores at all. Effect sizes are reported in standard deviation unit so that comparisons of effect sizes can be made across treatments and outcomes that are measured in different units.

[10] Aucejo, Esteban M., and Teresa Foy Romano. (2013). Assessing the effect of school days and absences on test score performance. CFP Discussion Paper No. 1302.

Indicator	Outcome Measure	Grade Level	Exemplary Finding Summary, Effect Sizes, etc.	ESSA Level of Research Evidence (See text above for details)	Strength of Causal or Correlational Relationship[8] (1 is strong-est, 4 is weakest)
Teacher Attendance	Academic achievement	K-12	Promising: missing 10 additional days[11] leads to 1.7%-3.2% SD decrease in math *per student.*[12]	2	2
Student Suspensions	Academic achievement, high school persistence	K-12	Promising: negative correlation with academic achievement – 49% of students who enter high school with 3 suspensions eventually drop out.[13]	3	2
School Climate[14]	Academic Achievement	K-12	Strong: school climate is positively correlated with many improved academic and behavioral outcomes.[15]	3	1

[11] Absences are defined here as teachers taking sick or personal leave. We acknowledge that teachers are contractually allowed to take a certain number of sick days and certainly do not advocate for a policy in which this is not the case. However, it should be noted that school districts could give teachers financial incentives for each day of contractually allowed leave not taken.

[12] Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2009). Are teacher absences worth worrying about in the United States?. Education, 4(2), 115-149.

[13] Balfanz, R., & Fox, J. (2015). Sent home and put off-track: The antecedents, disproportionalities, and consequences of being suspended in the ninth grade. Journal of Applied Research on Children: Informing Policy for Children at Risk, 5(2), 13.

[14] While there is a large research literature on school climate, and some of it strong, there is still not full agreement as to the components that get counted. Chiefs choosing to go this route should be advised that putting together a bundle of indicators through survey and observational instruments is a complex operation at all levels.

[15] Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. Review of Educational Research, 83(3), 357-385. . A Review of School Climate Research (PDF Download Available). Retrieved here.

Indicator	Outcome Measure	Grade Level	Exemplary Finding Summary, Effect Sizes, etc.	ESSA Level of Research Evidence (See text above for details)	Strength of Causal or Correlational Relationship[8] (1 is strong-est, 4 is weakest)
Student Retention	Academic achievement, high school persistence	K-12	Unclear for use in an accountability system: recent quasi-experimental research finds that retention has initially positive effects on academic achievement, but this effect fades after 6 years. These researchers also find retention in 3rd grade has no effect on the probability of high school graduation.[16]	2	3
Student Engagement	Academic achievement, high school persistence	K-12	Strong: student engagement is positively related to academic achievement and negatively related to the probability of dropping out of high school.	3	1

[16] Schwerdt, G., West, M. R., & Winters, M. A. (2015). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida (No. w21509). National Bureau of Economic Research. Retrieved here.

Indicator	Outcome Measure	Grade Level	Exemplary Finding Summary, Effect Sizes, etc.	ESSA Level of Research Evidence (See text above for details)	Strength of Causal or Correlational Relationship[8] (1 is strong-est, 4 is weakest)
Non-Cognitive Skills	Academic achievement	K-12	Unclear: it is not clear how these could be validly used in an accountability system. Which measures or skills to target and how to measure them have not emerged from the literature.[17] For example, some skills that are negatively associated with schooling outcomes are positively associated with increased wages in adulthood.[18]	3	1

[17] Borghans, Lex, Angela L. Duckworth, James J. Heckman, and Bas ter Weel. (2008). "The Economics and Psychology of Personality Traits." Journal of Human Resources, vol. 43, no. 4, 972–1059.
Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters assessing personal qualities other than cognitive ability for educational purposes. Educational Researcher, 44(4), 237-251.
[18] Papageorge, N. W., Ronda, V., & Zheng, Y. (2016). The Economic Value of Breaking Bad: Misbehavior, Schooling and the Labor Market. Draft (April 8, 2016).

Indicator	Outcome Measure	Grade Level	Exemplary Finding Summary, Effect Sizes, etc.	ESSA Level of Research Evidence (See text above for details)	Strength of Causal or Correlational Relationship[8] (1 is strongest, 4 is weakest)
Early Drop-Out Indicators:[19] -Attendance of 80% or less -Failing a math course -Failing an English course -Receiving an out-of-school suspension	High school persistence	6th grade	Strong: One study identified 60% of students in 6th grade who eventually drop out.[20]	3	1

[19] Using multiple measures together in practice is often useful. See: http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Accountability/Pages/default.aspx for an example of Georgia's multiple college readiness measures.
[20] Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. Educational Psychologist, 42(4), 223-235.

Indicator	Outcome Measure	Grade Level	Exemplary Finding Summary, Effect Sizes, etc.	ESSA Level of Research Evidence (See text above for details)	Strength of Causal or Correlational Relationship[8] (1 is strongest, 4 is weakest)
"On-Track" Drop-Out Indicators: -Enough credits to be promoted to 10th grade -No more than one F	High school persistence	9th grade	Strong: in one study, on-track students were 4 times more likely to graduate from high school than those who were off-track.[21]	3	1
Credit Accumulation	High school persistence	6-12	Strong: in one study, the number of failed courses correctly identified 80% of students who did not graduate.[22]	3	1

[21] Allensworth, E. M., & Easton, J. Q. (2005). The on-track indicator as a predictor of high school graduation. Chicago: Consortium on Chicago School Research.
[22] Allensworth, E. M., & Easton, J. Q. (2007). What matters for staying on track and graduating in Chicago Public High Schools. Chicago, IL: Consortium on Chicago school research.

Indicator	Outcome Measure	Grade Level	Exemplary Finding Summary, Effect Sizes, etc.	ESSA Level of Research Evidence (See text above for details)	Strength of Causal or Correlational Relationship[8] (1 is strong-est, 4 is weakest)
High School GPA	College readiness	9-12	Unclear for use in an accountability system: positive correlation with college GPA, persistence in 2nd year of college, college completion, and on-time completion.[23] There remain real concerns about grade inflation and lack of standardization between schools.[24]	3	3
International Baccalaureate Diploma (and Pre-Diploma) Program	Academic achievement, college readiness	9-12	Promising: 50% more likely to attend selective colleges, and 11% more likely to stay in college for at least 2 years.[25]	3	2
Advanced Placement Courses	Academic achievement, college readiness	9-12	Strong: students who achieve a college credit-granting score on their AP exams are 1-2% more likely to graduate with a Bachelor's degree within 4 years per exam.[26]	2	2

[23] Jackson, J., & Kurlaender, M. (2013). College readiness and college completion at broad access four-year institutions. American Behavioral Scientist, 0002764213515229.
[24] Woodruff, D. J., & Ziomek, R. L. (2004). High School Grade Inflation from 1991 to 2003. Research Report Series 2004-04. ACT Inc.
[25] Coca, V., Johnson, D., Kelley-Kemple, T., Roderick, M., Moeller, E., Williams, N., & Moragne, K. (2012). Working to My Potential: The Postsecondary Experiences of CPS Students in the International Baccalaureate Diploma Programme. Consortium on Chicago School Research.
[26] Note that a "college credit-granting score" depends on the college the student attends, although this is either a 3,4, or 5.

Indicator	Outcome Measure	Grade Level	Exemplary Finding Summary, Effect Sizes, etc.	ESSA Level of Research Evidence (See text above for details)	Strength of Causal or Correlational Relationship[8] (1 is strong-est, 4 is weakest)
Dual Enrollment	Academic achievement, college readiness	9-12	Strong: students are 6% more likely to attain any college degree (effect size of 0.14) and 5% more likely to attain a Bachelor's degree (effect size of 0.12) when they participated in a dual enrollment program.[27]	2	1

Student Attendance

Overall Finding: Unsurprisingly, student absences (both excused and unexcused) reduce academic achievement (Aucejo & Romano 2014; Goodman 2014; Gottfried 2009). Student absences also put students at risk for future truancy and increase their probability of dropping out of school (Alexander et al. 1997; Rumberger 1995). Goodman (2014) estimates that each absence caused by bad weather reduces math achievement by 5% of a standard deviation for elementary students and almost 4% for older students.[28] Other researchers have found more modest effect sizes: reducing students' absences by ten days per annum is estimated to increase math and reading test scores by 5.8% and 3% of a standard deviation respectively (Aucejo & Romano 2014). These effects are stronger for poor students and black or Hispanic students (Goodman, 2014; Aucejo & Romano 2014).

Possible Indicator: The average number of student absences per year or the change in the average number of student absences per year. This indicator can be disaggregated. Measures would likely need to be in place in order to ensure data accuracy.

Smith, J., Hurwitz, M., & Avery, C. (2015). Giving college credit where it is due: Advanced Placement exam scores and college outcomes (No. w21147). National Bureau of Economic Research. Retrieved here.

[27] An, B. P. (2012). The impact of dual enrollment on college degree attainment: Do low-SES students benefit? Educational Evaluation and Policy Analysis, 35, 57–75.

[28] These "older student" estimates are for grades 6th-8th and 10th grade.

Teacher Attendance

Overall Finding: Teacher absences reduce student learning (Miller et al., 2008; Clotfelter et al., 2009). Negative impacts can be both academic and emotional, and substitute teachers contribute significantly less to student learning (Herrmann et al. 2010). The overall effect sizes in research are at least moderate. One study estimates that 10 additional days of a teacher's absence diminishes fourth-graders' mathematics achievement by 3.2% of a standard deviation (Miller et al., 2008). Clotfelter et al. (2009) find statistically significant, but smaller, effect sizes (1.7% of a standard deviation decrease in math achievement and 0.9% percent standard deviation in reading). Even though these effect sizes appear small, they represent the average reduction *per student* in the class and therefore reflect much larger learning losses when aggregated over the class.

Interestingly, research also finds that unexpected teacher absences decrease student learning more than do anticipated teacher absences (Miller et al., 2008). Using data from North Carolina, researchers found that teachers serving low-income students were more likely to be absent than those teaching high-income students. Teachers of students in the lowest income quartile took almost one additional sick day a year on average when compared to teachers serving students in the highest-income quartile (Clotfelter et al., 2009).

Teacher absenteeism is a potentially promising data element, since it is already reported at the LEA level to the U.S. Department of Education for the Office for Civil Rights, though data quality would likely need to improve. Teacher attendance is currently used in New York state as part of required reporting for the state's School Improvement Grants. At least sixteen states currently collect this data, with six reporting it in state report cards or accountability systems (D'Agati & Szuberla, 2015).

Possible Indicator: The average number of teacher absences per year. This indicator cannot be easily disaggregated, although a weighted average could be calculated.[29] Measures would likely need to be in place in order to ensure data accuracy.

[29] The teacher's absences could be weighted by the percentage of students taught by that teacher within a given subgroup.

Student Suspensions

Overall Finding: School suspensions are correlated with a host of negative outcomes, but research has yet to establish that this relationship is causal. Specifically, students who are suspended from school have lower academic achievement (Rumberger 1995; Advancement Project/Civil Rights Project 2000; Mendez 2003; Rausch & Skiba, 2005 Morris & Perry 2016) and are more likely to be retained in a grade and to drop out of school (Mendez, 2003; Gregory, Skiba, & Noguera, 2010). Research also finds that schools with higher levels of suspensions are associated with the lower achievement of non-suspended students (Perry & Morris, 2014). Males, students of color, those who are in special education, and/or students who qualify for free and reduced lunch, are much more likely to be suspended. Suspension in an earlier grade is positively correlated with suspension in later grades (Mendez, 2003; Mendez & Knoff, 2003; Gregory, Skiba, & Noguera, 2010).

Possible Indicator: The average number or the change in the average number of student suspensions per year. This indicator can be disaggregated. Measures would likely need to be in place in order to ensure data accuracy and that students are not being suspended in order to move this indicator. States would also have to ensure that n-sizes were large enough to use in a valid and reliable manner.

School Climate

Overall Finding: Durham et al. define "school climate" as how students, teachers,[30] and parents view their school. School climate is positively correlated with social, emotional, and academic outcomes (Thapa et al., 2013). There is a growing body of work supporting the importance of school climate in promoting academic achievement, school safety, dropout prevention, teacher retention, healthy social interactions, and well-being (Cohen, 2010; Dynarski et al., 2008). Furthermore, research has found that students across all grades in K-12 education are more likely to achieve higher test scores in schools with positive school climates (MacNeil et al., 2009).

Researchers tend to identify five areas of school climate: safety, relationships, teaching and learning,

[30] Note that "teacher engagement" can be considered one aspect of school climate. The ESSA gives teacher engagement as an example of a potential ISQ, however we have found very little research basis or valid measures that specifically isolate teacher engagement.

institutional environment, and the school improvement process (Thapa et al., 2013). Positive school climate is positively associated with many student outcomes, including better student mental and physical health, decreased student absences, and lower suspension rates (Thapa et al., 2013). A supportive school climate is also associated with increased teacher effectiveness (Kraft & Papay, 2014).

Possible Indicator: Numerous validated school-climate tools currently exist[31] including the Comprehensive School Climate Inventory, which surveys school personnel, students, and parents to measure safety; teaching and learning; interpersonal relationships; and the institutional environment (Guo et al., 2011). This indicator cannot be easily disaggregated. Parent and student surveys could be combined with weighted teacher scores[32] to arrive at a disaggregated measure, although it is unclear if the school-climate tools are valid for this purpose. Measures would likely need to be in place in order to ensure data accuracy as survey responses can be manipulated.

Student Retention

Overall Finding: Early research found that retaining students is associated with many negative outcomes including decreased socio-emotional outcomes, completion of fewer years of schooling, and decreased student learning (for example see: McCoy and Reynolds, 1999; Jimerson, 1999; Jimerson, 2001; Jimerson et al., 2002). However, more recent research concludes that higher-quality studies[33] found less negative effects of grade retention on student outcomes (Allen et al., 2009). Indeed a recent quasi-experimental study of 3rd graders in Florida found that retaining students initially produces positive effects on achievement, although these effects fade out after 6 years. Further, in contrast to previous findings, this study finds that 3rd grade retention has no effect on the probability that students will graduate from high school (Schwerdt et al., 2015).

Possible Indicator: The percentage of students retained per year. This indicator can be disaggregated.

---

[31] Other school climate surveys that have been validated include: Alliance for the Study of School Climate–School Climate Assessment Inventory (Schindler et al., 2003); 5Essentials (Bryk et al., 2010); Inventory of School Climate-Teacher (Brand et al, 2008); and School-Climate Inventory-Revised (Butler and Rakow, 1995).

[32] i.e. weighted by the percentage of students within a given subgroup the teacher teaches

[33] The researchers defined research quality in several ways. For example, they defined "comparison group quality" as high if the average pre-retention performance of the promoted group (on a ability measure or other initial outcome measure) was equivalent to the average pre-retention performance of the retained group. "Statistical control quality" as was given a high rating the researchers statistically controlled for a pre-retention measure of the outcome variable.

Student Engagement

Overall Finding: Student engagement is associated with both academic achievement and high school dropout rates. Engaged students are less likely to drop out of high school (Fredricks et al., 2004; Rumberger & Rotermund, 2012). Students that are engaged in school are also more likely to both earn higher grades and perform better on standardized tests (Marks, 2000; Fredricks et al., 2004; Wigfield et al., 2015). Empirical analysis has shown that students feel most engaged in elementary school, and that engagement typically declines with age, although this decline also depends on external circumstances (Wigfield et al., 2015). For example, the rate that student engagement declines can be faster in high-poverty schools that are low-performing (Yazzie-Mintz, 2007).

Possible Indicator: There are numerous validated measures of student engagement (Fredricks et al., 2011). Engagement can be measured using student surveys, teacher reports, or observational measures. Student surveys can be disaggregated, although are potentially easy to manipulate. Observational measures are harder to manipulate, but present challenges to disaggregate or are costly.[34]

Non-Cognitive Skills

Overall Finding: There is currently not sufficient evidence to suggest that these measures are ready for adoption in a state accountability framework. While these measures may be helpful to examine as part of a diagnostic review process or of school-level reporting, there can be unintended consequences for adopting measures into an accountability system that lack validity or effective measurement tools (Hess et al., 2009).

Interest in this topic has been generated by emerging research that shows some positive impacts of strong non-cognitive skills upon academic achievement. Egalite et al. define "non-cognitive skills" as "a set of behaviors, attitudes, and strategies that have been shown to be associated with individual success. This incorporates constructs such as optimism, resilience, adaptability, and conscientiousness." However, as this very recent research study concludes, "Despite popular interest

[34] Note that some observational measures observe individual students, and therefore might be costly to collect. Other observational measures observe the classroom, which might be difficult to disaggregate.

in this topic, our knowledge of how to measure these skills is still underdeveloped." (Egalite et al., 2015)

Recent research builds on prior studies that indicate the limitations of traditional testing to predict future employment and income success. The challenges remain that the behaviors used to measure particular non-cognitive traits can be influenced by incentives and by other traits, and that we know little about the possible interference effect[35] of measuring non-cognitive skills in a high-stakes environment. Further research suggests that, "indicators of adolescent skills have strong associations with educational attainment, but not subsequent labor market outcomes, and illustrate some problems in interpreting apparent skill gaps across demographic groups (Lundberg 2015)."

Additionally, questions have been raised about the very possibility of measuring non-cognitive skills as variables that are independent of academic behavior when: "Academic behaviors are extremely important for achievement…. virtually all other non-cognitive factors work *through* academic behaviors to affect performance." Academic behaviors – attending class, doing homework, and studying – are easily measurable in the first case, as per "attendance" above, or cannot be easily measured at all (Farrington et al., 2012).

It is telling that in a recent opinion piece, Angela Duckworth, the researcher most associated with non-cognitive skills such as grit and perseverance, strongly advocates not using such traits in formal accountability measures (see this *New York Times* article and this article written by Angela Duckworth).

Possible Indicator: The average (or weighted) score from student surveys that measures a specific non-cognitive skill. This indicator can be disaggregated. Note that this indicator can be easily manipulated and therefore measures would need to be in place in order to ensure data accuracy.

Early Warning Indicators

Overall Finding: Well-known research from Johns Hopkins offers strong evidence that measuring a small number of variables at the middle-school level offers critical, predictive information about high school completion Ruth Neild, Robert Balfanz, and Lisa Herzog, "An Early Warning System," *Educational Leadership*, October 2007, http://new.every1graduates.org/wp-

[35] For example, through the directions that students are given before completing their surveys.

content/uploads/2012/03/Early_Warning_System_Neild_Balfanz_Herzog.pdf.. In particular, "a 6th grader with even one of the following four signals had at least a three in four chance of dropping out of high school: A final grade of F in mathematics; a final grade of F in English; attendance below 80 percent for the year; or a final "unsatisfactory" behavior mark in at least one class." A more recent analysis of 8th and 9th graders confirms that a system of similar indicators robustly predicts subsequent high school drop-out (Henry et al., 2012). Balfanz's work is also used at the start of high school: The state of Louisiana has created a Dropout Credit Accumulation Index that measures schools by the numbers of credits accumulated through 9th grade (Louisiana's High School Student Planning Guidebook, 2016-17).

Possible Indicator: The percentage of students who have one or more indicator per year. This indicator can be disaggregated. Measures would likely need to be in place in order to ensure data accuracy. States would also have to ensure that n-sizes were large enough to use in a valid and reliable manner.

Credit Accumulation

Overall Finding: Multiple indicators, including high absenteeism, poor behavior, negative relationships with teachers, and poor grades, predict whether a student will drop out of high school. Studies have shown that students who fail classes and/or fail to earn enough credits for promotion to the next grade are less likely to graduate from high school (Bowers & Sprout 2012; Jerald 2006). However, the most accurate predictive measures seem to be students' attitudes about school, students' behavior, and credit accumulation.

Credit accumulation is therefore especially informative when combined with other measures as part of an early warning indicator for high school dropout. For example, a study of Chicago Public Schools found that 9th students who were "on-track" (e.g., had accumulated enough credits to be promoted to 10th grade and had no more than one F in English, math, science, or social studies) were roughly 4 times more likely to graduate from high school on time than those who were not (i.e., 20% of off-track students graduated in 4 years in comparison to 81% of on-track students) (Allensworth and Easton, 2005). Using the "on-track" indicator from 9th grade correctly identified whether students graduated from high school 80% of the time (Allensworth and Easton, 2007). Failing courses in 9th grade reduces the probability a student will graduate from high school.

Possible Indicator: The percentage of students who have one or more "on-track" indicator per year. This indicator can be disaggregated. Measures would likely need to be in place in order to ensure data accuracy.

Grade Point Averages (GPA)

Overall Finding: Research finds that high school GPAs are more predictive of first-year college GPAs than are tests (Sawyer, 2010) and that they provide useful information about students' non-cognitive abilities (Bowers, 2010). However, grade inflation is a large problem with using GPA as a college readiness measure: Researchers have documented the rise of mean GPAs, despite the fact that other measures of college readiness have not risen (Woodruff, 2004). Furthermore, there is a lack of standardization across classrooms and schools that renders comparability problematic.

Possible Indicator: The average GPA in a school per year. This indicator can be disaggregated. This indicator can be easily manipulated and measures would need to be in place in order to ensure data accuracy.

Course Completion (includes IB, AP, and DE)

Overall Finding: In analyzing student transcripts, Adelman (2006) found that completing a challenging high school curriculum is the greatest pre-collegiate indicator of Bachelor's degree completion; this impact is even greater for black and Hispanic students than for white students. Twenty-five states currently require all school districts to offer IB, AP, and/or dual enrollment.

Possible Indicator: The percentage of students who participated in a given number of challenging courses such as IB, AP, or DE courses in the school year. This indicator can be disaggregated. Measures would likely need to be in place in order to ensure data accuracy.

International Baccalaureate (IB)

Overall Finding: While there is not a large body of research on the effects of the IB program on student learning, initial findings are promising. Public high-school students in Chicago who successfully completed the International Baccalaureate Diploma Program (IBDP) were 50% more likely to attend

selective colleges and significantly more likely to stay in college for at least two years, than their matched peers. Qualitative analysis indicated that the students' experiences in the IBDP taught them specific skills and behaviors required in college (Coca et al., 2012). (Approximately half of the students who began the IB program in 9th grade completed the IB Diploma – there were no negative academic impacts for those who did not). A further study showed that the IBDP also increased students' probability of earning a B average or better in 11th grade coursework, with the greatest performance increase in mathematics (Cortes et al., 2013). Participation in IBDP also increased ACT scores (Saavedra, 2011).

Possible Indicator: The percentage of students who participated in IB courses and successfully completed their IBDP. This indicator can be disaggregated.

Advanced Placement (AP) Coursework

Overall Finding: There is significant research on the effects of AP courses on student learning, although the results are mixed as to whether taking an AP course alone is a sufficient measure of college readiness, or if both taking the course and achieving a certain score on the AP test is a superior measure.

For example, in-house researchers from the College Board found a positive relationship between AP courses and many college outcomes. They found that students who took one or more AP courses and exams had higher college GPAs, earned more credit hours, and were more likely to graduate in 4 years or less (Hargrove et al., 2008). College Board in-house researchers also found that students who had higher AP exam scores were more likely to have higher grades in the first year of college (Shaw et al., 2013). These findings suggest that taking an AP course alone is a sufficient college-readiness indicator.

However, independent research finds that both taking an AP course and passing the AP exam is a superior measure of college readiness. For example, Dougherty et al. (2005) conclude that the percentage of students who both take and pass AP exams is the best AP-related measure that a school is preparing students to graduate from college. Further, Smith et al. (2015) use quasi-experimental methods to show a causal relationship between AP exam scores and college graduation. Specifically, the researchers show that earning a college credit-granting score on the AP exam increases the probability that a student will receive a Bachelor's degree within four years by 1-2 percentage points per exam.

Possible Indicator: The percentage of students who both participated in an AP course and earned a college credit-granting score[36] on the AP exam. This indicator can be disaggregated. This indicator would be hard to manipulate.

Dual Enrollment (DE)/Early College

Overall Finding: While research is still limited in this area, one study does find a causal relationship between DE programs and college outcomes. Specifically, An (2012) finds that students in DE programs are 6% more likely to attain any college degree and 5% more likely to attain a Bachelor's degree. Other researchers have found that participation in DE programs is positively correlated with a higher college GPA, as well as higher college graduation rates (Karp et al., 2007; Speroni, 2011).

Possible Indicator: The percentage of students who participated in a DE course and earned a passing grade in that course. This indicator can be disaggregated.

[36] Since this score depends on the college the student ultimately attends, states and districts could define this in various ways, such as choosing the lowest score needed for college credit (i.e. a 3), or the score needed on average (rounded to the nearest integer) for the previous years' graduating class, for example.

SOURCES

Main Text

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237-251.

Hansen, M. (2016). New school accountability measures must be careful not to reinforce student poverty. Brown Center Chalkboard. Brookings. Retrieved here.

Kemple, J. J., & Snipes, J. C. (2000). Career Academies: Impacts on students' engagement and performance in high school. New York, NY: MDRC.

United States Department of Education. Elementary and Secondary Education Act of 1965, as Amended by the Every Student Succeeds Act-Accountability and State Plans. The Federal Register. November 29, 2016. Retrieved here.

Student Attendance

Alexander, KL, DR Entwisle, and CS Horsey. (1997). From First Grade Forward: Early Foundations of High School Dropout. Sociology of Education, 70(2), 87-107. Retrieved here.

Aucejo, Esteban M., and Teresa Foy Romano. (2013). Assessing the effect of school days and absences on test score performance. CFP Discussion Paper No. 1302. Retrieved here.

Ensminger, ME, and AL Slusarcick. (1992). Paths to High School Graduation or Dropout: A Longitudinal Study of a First-Grade Cohort. Sociology of Education, 65(2), 95-113. Retrieved here.

Goodman, J. (2014). Flaking out: Student absences and snow days as disruptions of instructional time. National Bureau of Economic Research Working Paper No. w20221. Retrieved here.

Gottfried, MA. (2009). Excused versus Unexcused: How Student Absences in Elementary

School Affect Academic Achievement. Educational Evaluation and Policy Analysis, 31(4), 392-415. Retrieved here.

Rumberger, RW. (1995). Dropping out of Middle School: A Multilevel Analysis of Students and Schools. American Educational Research Journal, 32(3), 583-615. Retrieved here.

Teacher Attendance

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2009). Are teacher absences worth worrying about in the United States?. *Education*, *4*(2), 115-149. Retrieved here.

D'Agati, J. & Szuberla, C. (2015). "Collection of Teacher-Level Attendance Data". Albany, New York: New York State Department of Education. Retrieved here.

Herrmann, Mariesa, and Jonah Rockoff. "Worker Absence and Productivity: Evidence from Teaching." The National Bureau of Economic Research Working Paper 16524 (2010). www.nber.org. Retrieved here.

Miller, R. T., Murnane, R. J., & Willett, J. B. (2008). Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. *Educational Evaluation and Policy Analysis*, *30*(2), 181-200. Retrieved here.

Student Suspensions

Advancement Project/Civil Rights Project. (2000) "Opportunities Suspended: The Devastating Consequences of Zero Tolerance and School Discipline." A report from the National Summit of Zero Tolerance: Washington, DC. Retrieved here.

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The Achievement Gap and the Discipline Gap Two Sides of the Same Coin?. *Educational Researcher*, *39*(1), 59-68. Retrieved here.

Mendez, LMR. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. New Directions for Youth Development, 99, 17-33. Retrieved here.

Mendez, LMR, and HM Knoff. (2003). Who gets suspended from school and why: A demographic analysis of schools and disciplinary infractions in a large school district.

Education and Treatment of Children, 26(1), 30-51. Retrieved here.

Mendez, LMR, HM Knoff, and JM Ferron. (2002). School demographic variables and out-of-school suspension rates: A quantitative and qualitative analysis of a large, ethnically diverse school district. Psychology in the School, 39(3), 259-277. Retrieved here.

Morris, E. W., and B. L. Perry, (2016). "The Punishment Gap: School Suspension and Racial Disparities in Achievement." Social Problems, vol 63, no. 1, pp. 68-86. Retrieved here.

Perry, B. L., & Morris, E. W. (2014). Suspending Progress Collateral Consequences of Exclusionary Punishment in Public Schools. *American Sociological Review*, 0003122414556308. Retrieved here.

Rausch, M. Karega and R. J. Skiba. (2005) "The Academic Cost of Discipline: The Relationship Between Suspension/Expulsion and School Achievement." Center for Evaluation and Education Policy, School of Education, Indiana University. Retrieved here.

School Climate

Brand, S., Felner, R., Seitsinger, A., Burns, A., & Bolton, N. (2008). A large scale study of the assessment of the social environment of middle and secondary schools: The validity and utility of teachers' ratings of school climate, cultural pluralism, and safety problems for understanding school effects and school improvement. Journal of School Psychology, 46, 507–535. Retrieved here.

Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). Organizing schools for improvement: Lessons from Chicago. University of Chicago Press.

Butler, D. E., & Rakow, J. (1995). Sample Tennessee school climate profile. Memphis, TN: University of Memphis, Center for Research in Educational Policy. Retrieved here.

Cohen, J. (2010). School Climate Research Summary. Center for Social and Emotional Education. Found here.

Guo, P., Choe, J., & Higgins-D'Alessandro, A. (2011). Report of construct validity and internal consistency findings for the comprehensive school climate inventory. Retrieved here.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, *36*(4), 476-500. Retrieved here.

MacNeil, A. J., Prater, D. L., & Busch, S. (2009). The effects of school culture and climate on student achievement. *International Journal of Leadership in Education*, *12*(1), 73-84.

Shindler, J., Taylor, C., Cadenas, H., & Jones, A. (2003). Sharing the Data along with the Responsibility: Examining an Analytic Scale-Based Model for Assessing School Climate. Retrieved here.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357-385. Retrieved here.

Student Retention

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, *31*(4), 480-499. Retrieved here.

Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. Journal of School Psychology, 37(3):243–272.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School psychology review*,*30*(3), 420. Retrieved here.

Jimerson, S. R., Anderson, G. E., and Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. Psychology in the Schools, 39(4):441–457.

McCoy, A. R. and Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. Journal of School Psychology, 37(3):273–298.

Schwerdt, G., West, M. R., & Winters, M. A. (2015). *The effects of test-based retention on*

student outcomes over time: Regression discontinuity evidence from Florida (No. w21509). National Bureau of Economic Research. Retrieved here.

Student Engagement

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, *74*(1), 59-109. Retrieved here.

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., and Mooney, K. (2011). Measuring student engagement in upper elementary through high school: a description of 21 instruments. (Issues & Answers Report, REL 2011–No. 098). Washington, DC: U.S.

Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved here.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. American educational research journal, 37(1), 153-184. Retrieved here.

Rumberger, R. W., & Rotermund, S. (2012). The relationship between engagement and high school dropout. In Handbook of research on student engagement (pp. 491-513). Springer US. Retrieved here.

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2015). Development of achievement motivation and engagement. Handbook of child psychology and developmental science. Retrieved here.

Yazzie-Mintz, E. (2007). Voices of Students on Engagement: A Report on the 2006 High School Survey of Student Engagement. Center for Evaluation and Education Policy, Indiana University. Retrieved here.

Non-Cognitive Skills

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237-251.

Farrington, C.A., M. Roderick, E. Allensworth, J. Nagaoka, T.S. Keyes, D. W. Johnson, N. O. Beechum. (2012). *Teaching Adolescents to Become Learners: The Role of Non-cognitive Factors in Shaping School Performance: a Critical Literature Review*. Chicago: University of Chicago Consortium on Chicago School Research.

Hess, F.M. & Fullerton, J. (2009). *Balanced Scorecards and Management Data*. Center for Education Policy Research, Harvard University. Retrieved here.

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, *19*(4), 451–464.

Lundberg, C. October 16-17, (2015). Non-Cognitive Skills as Human Capital. *NBER/CRIW Conference on Education, Skills, and Technical Change*.

West, Martin. March 30, (2016). "Should Non-Cognitive Skills Be Included in School Accountability Systems?" *EducationNext.* Retrieved here.

Early Warning Indicators

Louisiana's High School Student Planning Guidebook: A Path to Prosperity for Every Student (2016-17).

Henry, K. L., Knight, K. E., & Thornberry, T. P. (2012). School disengagement as a predictor of dropout, delinquency, and problem substance use during adolescence and early adulthood. *Journal of youth and adolescence*, *41*(2), 156-166. Retrieved here.

Neild, R., Balfanz., R., Herzog, L. (October 2007). "An Early Warning System." Educational Leadership.

Credit Accumulation

Allensworth, E. M., & Easton, J. Q. (2005). The on-track indicator as a predictor of high school graduation. Chicago: Consortium on Chicago School Research. Retrieved here.

Allensworth, E. M., & Easton, J. Q. (2007). What matters for staying on track and graduating in Chicago Public High Schools. *Chicago, IL: Consortium on Chicago school research.* Retrieved here.

Bowers, A. J., & Sprott, R. (2012). Examining the multiple trajectories associated with dropping out of high school: A growth mixture model analysis. *The Journal of Educational Research*, *105*(3), 176-195. Retrieved here.

Jerald, C. D. (2006). Dropping Out Is Hard to Do. Issue Brief. *Center for Comprehensive School Reform and Improvement*. Retrieved here.

Grade Point Average (GPA)

Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation*, *17*(3), 141-159. Retrieved here.

Sawyer, R. (2010). Usefulness of High School Average and ACT Scores in Making College Admission Decisions. ACT Research Report Series 2010-2.*ACT, Inc.* Retrieved here.

Woodruff, D. J., & Ziomek, R. L. (2004). High School Grade Inflation from 1991 to 2003. Research Report Series 2004-04. *ACT Inc.*. Retrieved here.

Course Completion

Adelman, C. (2006). The toolbox revisited: Paths to degree completion from high school through college. Washington, DC: U.S. Department of Education. Retrieved here.

International Baccalaureate

Coca, V., Johnson, D., Kelley-Kemple, T., Roderick, M., Moeller, E., Williams, N., & Moragne, K. (2012). Working to My Potential: The Postsecondary Experiences of CPS Students in the International Baccalaureate Diploma Programme. Consortium on Chicago School Research.

Retrieved here.

Cortes, K. E., Moussa, W. S., & Weinstein, J. M. (2013). Educating bright students in urban schools. *Economics of Education Review*, *37*, 286-297. Retrieved here.

Saavedra, A. R. (April 2011). The academic impact of enrollment in International Baccalaureate Diploma Programs: A case study of Chicago Public Schools. In *annual conference of the American Educational Research Association, New Orleans, LA*. Retrieved here.

Advanced Placement Coursework

Dougherty, C., Mellor, L., & Jian, S. (2006). The Relationship between Advanced Placement and College Graduation. 2005 AP Study Series, Report 1. *National Center for Educational Accountability*.

Hargrove, L., Godin, D., & Dodd, B. G. (2008). College outcomes comparisons by AP and non-AP high school experiences. Retrieved here.

Jeong, D. W. (2009). Student Participation and Performance on Advanced Placement Exams: Do State-Sponsored Incentives Make a Difference?.*Educational Evaluation and Policy Analysis*, *31*(4), 346-366. Retrieved here.

Klopfenstein, K., & Thomas, M. K. (2009). The link between advanced placement experience and early college success. *Southern Economic Journal*, 873-891. Retrieved here.

Shaw, E. J., Marini, J. P., & Mattern, K. D. (2013). Exploring the utility of Advanced Placement participation and performance in college admission decisions. *Educational and Psychological Measurement*, *73*(2), 229-253. Retrieved here.

Smith, J., Hurwitz, M., & Avery, C. (2015). *Giving college credit where it is due: Advanced Placement exam scores and college outcomes* (No. w21147). National Bureau of Economic Research. Retrieved here.

Dual Enrollment/Early College

An, B. P. (2012). The Impact of Dual Enrollment on College Degree Attainment Do Low-SES Students Benefit?. *Educational Evaluation and Policy Analysis*, Retrieved here.

Berger, A., Turk-Bicakci, L., Garet, M., Knudson J., Hoshen, G. (2014). Early College, Continued Success: Early College High School Initiative Impact Study. American Institutes for Research.

Karp, M. M., Calcagno, J. C., Hughes, C. L., Jeong, D. W., & Bailey, T. (2007). The postsecondary achievement of participants in Dual Enrollment: An analysis of student outcomes in two states. St. Paul, MN: University of Minnesota, National Research Center for Career and Technical Education. Retrieved here.

Speroni, C. (2011). Determinants of Students' Success: The Role of Advanced Placement and Dual Enrollment Programs. An NCPR Working Paper. *National Center for Postsecondary Research*. Retrieved here.