February 2017

CCSSO
Council of Chief State School Officers

Innovative Approaches and Measurement Considerations for the Selection of the School Quality and Student Success Indicator under ESSA

Katie Buckley, Transforming Education
with
Ajit Gopalakrishnan, Connecticut Department of Education
Evan Kramer, Tennessee Department of Education
Andy Whisman, West Virginia Department of Education

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nonpartisan, nationwide nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. The Council seeks member consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS
Melody Schopp (South Dakota), President
Chris Minnich, Executive Director

One Massachusetts Avenue, NW, Suite 700 • Washington, DC 20001-1431
Phone (202) 336-7000 • Fax (202) 408-8072 • www.ccsso.org
Copyright © 2017 by the Council of Chief State School Officers, Washington, DC
All rights reserved.

ABSTRACT

This paper, written in collaboration with the Council of Chief State School Officers (CCSSO), provides a deep and nuanced examination of how states might respond to the student success or school quality indicator accountability provision (i.e., the so-called "5th indicator") under the Every Student Succeeds Act (ESSA). Conversations with state personnel reveal state education agencies (SEAs) are evaluating the range of choices for the school quality or student success indicator. Specifically, states are considering if indicators already being measured for state reporting or district accountability are suitable, which offers an advantage of reducing the implementation burden. Alternatively, SEAS are considering implementing new indicators to encourage an expanded definition of student and school success going beyond the requirements of the No Child Left Behind Act (NCLB) of 2001.

Leveraging the experience and leadership of states through CCSSO, we provide guidance on indicators, measures, and pathways being considered by state leaders to push the boundaries of what SEAs previously included in their accountability systems and explore what can be implemented inexpensively at scale. We focus on three indicators: (1) College and Career Readiness, (2) School Climate, and (3) Student Social-Emotional Mindsets and Skills and discuss the ways each indicator can be operationalized within an accountability system. We then provide brief case studies of three states seeking innovative pathways to comply with ESSA's requirement of including at least one other indicator of school quality or student success in the state accountability system. Finally, we discuss key measurement concerns for states to be attuned to as they plan for, implement, and evaluate their updated systems.

CONTENTS

Introduction 3

Selecting Indicators and Measures of School Quality and Student Success 4

College and Career Readiness 6

School Climate 9

Student Social-Emotional Mindsets and Skills 11

Pathways to Innovation: Multiple Indicators, Multiple Measures, and Adaptive Measures 14

Multiple Indicators 14

Box 1: Connecticut's Approach to Multiple Indicators under ESSA 16

Multiple Measures 17

Box 2: Tennessee's Consideration of a Multiple-Measures Index 18

Adaptive Measures 20

Box 3: Incorporating Drop-out Prevention into School Accountability: West Virginia's Early Warning System 21

Measurement Considerations 22

Identifying Where to Focus Measurement Activity 22

Aligning Validation Efforts to a Theory of Action 23

Determining How to Calculate the Metric 24

Combining Multiple Indicators into a Single Composite Accountability Score 25

System Evaluation 26

Conclusion 26

INTRODUCTION

With the passage of the Every Student Succeeds Act (ESSA), states were given wide authority to construct a school accountability model that can best advance college- and career-ready outcomes in their unique contexts. One of the most noteworthy elements of the revised statute was the creation of at least one additional indicator of school quality or student success. The school quality or student success indicator is sometimes referred to as the "5th indicator," as it is required in conjunction with four other indicators: academic achievement, student growth, graduation rate, and progress in achieving English language proficiency.

To ensure the school quality or student success indicator contributes to a state's system of school and district accountability, ESSA outlines several requirements. The indicator (or multiple indicators) must allow for "meaningful differentiation in school performance" and be "valid, reliable, and comparable across the state." The chosen indicator(s) must be the same across schools, but can vary by grade span. And the indicator(s) must be able to be disaggregated across student subgroups. While ESSA does not dictate how much weight the 5th indicator should be given, it does require that the weight of each of the first four be "substantial" and in aggregate "much greater" than the 5th indicator.

ESSA accountability provisions suggest the following as examples of a 5th indicator: "(I) student engagement; (II) educator engagement; (III) student access to and completion of advanced coursework; (IV) postsecondary readiness; (V) school climate and safety; and (VI) any other indicator the State chooses that meets the requirements of this clause."[1] Many education experts interpret this provision to mean that a host of measures, ranging from Advanced Placement (AP) test scores to culture and climate surveys, would satisfy the requirements of the law. However, the law requires the additional indicator to have a research base linked to improving student academic achievement. Thus, while the 5th indicator certainly affords an opportunity to move beyond test scores, it appears to constrain the selection of measures based on an association with student academic outcomes.

ESSA opens the door for accountability systems that focus on more than students' academic achievement. Therefore, the indicator of school quality or student success can be viewed as an opportunity for innovation. In particular, it is an opportunity for states to attend to and support those factors that go beyond academic test scores, as long as they have been concretely linked to positive student achievement. However, it also presents a challenge for states as they seek to incorporate additional measures into the current frameworks they use to hold schools and districts accountable. In this paper, we address some of the challenges associated with the choice of the 5th indicator, focusing on measurement considerations involved in operationalizing the chosen indicator.

1 Every Child Succeeds Act, Sec. 1111(c)(4)(B)(v).

SELECTING INDICATORS AND MEASURES OF SCHOOL QUALITY AND STUDENT SUCCESS

By collaborating with states through CCSSO, we have discovered a growing interest among education leaders to incorporate the following indicators into state accountability frameworks:

- College and Career Readiness: Are students graduating with the preparation needed for college or career without the need for remediation or learning support courses?
- School Climate: Do students and staff feel emotionally and physically safe in the school environment?
- Student Social-Emotional Mindsets and Skills: Are schools helping to develop students' social-emotional skills, such as self-management and social awareness?

Because these are quite broad, there is a good deal of flexibility in how a state chooses to measure the indicator. For the purposes of this paper, we define *indicator* as the broad construct of student success or school quality, we define *measure* as the way in which the indicator is assessed, and we define *metric* as the way in which the measure is scored for inclusion in the accountability system. For example, student social-emotional skills would be the indicator, student surveys would be the measure, and the percentage of positive responses would be the metric. In some cases, there may be overlap between the indicator and the measure, such as with "participation in advanced coursework." As we discuss in greater detail below, an indicator such as college and career readiness (CCR) incorporates many different facets of student success and school quality, and can therefore be assessed with a variety of measures ranging from participation in advanced coursework to aspects of the school culture to student social-emotional skills. As such, while choosing an indicator may narrow the pool of available measures, states will still need to focus considerable attention on exactly how to operationalize the indicator. Moreover, it is unlikely that one measure will suffice across all grade spans and schools; therefore, states will need to think through the choice of measures across grade spans (i.e., from elementary to middle to high schools) as well as within grade spans across schools (i.e., from vocational to traditional high schools).

When selecting a measure for inclusion in an accountability system, the following must be taken into account: actions the state wants to incentivize; goals of the accountability system; unintended consequences associated with a particular measure; and actionable results for educators. Each state department of education (SEA) that has contributed to this paper – Connecticut, Tennessee, and West Virginia – has explicitly articulated guiding principles for choosing the measures for their 5th indicator. For example, and to support states in their individual planning processes, the Tennessee Department of Education uses the following questions to guide their selection of the measure:

1. **Impact/validity/reliability**: Does the resulting metric demonstrate high-leverage impact on student outcomes? Does research support the measure as a valid and reliable indicator of school quality and/or student success?

2. **Buy-in**: Is there buy-in from district and school leaders?

3. **State vs. local role**: Is the state or local education agency better equipped to measure (and to inform practice to influence) this metric?

4. **Data quality**: Is the data quality sufficient to warrant inclusion in high-stakes evaluations?

5. **Differentiation**: Do the data vary sufficiently and meaningfully across schools and districts?

6. **Timeliness**: Does the measure allow for timely reporting?

While the guiding questions are not necessarily unique to Tennessee as many are drawn directly from the language of ESSA (e.g., validity, reliability, and differentiation), the prioritization that a state places on each of the guiding questions will likely shape which metrics are considered. For example, a state that prioritizes timeliness above all with the belief that accountability should drive immediate action on the part of districts and schools may not choose to incorporate lagged measures of postsecondary readiness.

A Chiefs for Change report argues the chosen measure requires not only a research base supporting its validity and reliability, but also research supporting (a) its use in a high-stakes framework and (b) the notion that the school is the correct vehicle to facilitate impact.[2] In terms of ensuring the measure is viable in a high-stakes framework, the authors note that states should proceed with caution when choosing an indicator "where adults have incentives and the ability to change student responses or produce simplistic solutions to complex educational challenges."[3] An example is that of suspension rates, which could be chosen as a measure of the school culture and climate. In a high-stakes framework, however, this might incentivize a school to simply stop suspending students for misbehavior without addressing the fundamental obstacles for at-risk students. Further, research should show that the chosen measure is something that the school actually has control over. For example, while the degree of student absenteeism is strongly predictive of academic outcomes, it has also been shown to correlate with socio-economic status. Therefore, it may be especially difficult for schools with large populations of children from a lower socio-economic background to show growth on this specific measure.

Another consideration is that the measure chosen for one grade span may not be appropriate for another grade span. For this reason, ESSA allows variation in the measure across grade spans. For example, in measuring college and career readiness it would clearly not make sense to measure the percentage of students with advanced coursework or career training at the elementary school level. Rather, the focus should be on measures that have been shown to predict later college and career readiness, such as on-track grade progression, access to and participation in gifted/accelerated courses, or student feelings of school belonging. Collecting predictive measures early on has the added benefit of targeting supports for students before it's too late to address the problem (i.e., before students drop out of school).

2 Chiefs for Change. (2016). *ESSA indicators of school quality and student success*. Author: Washington, DC. Retrieved December 26, 2016, from http://chiefsforchange.org/policy-paper/3244/.

3 Chiefs for Change. (2016). Retrieved December 27, 2016, from http://chiefsforchange.org/policy-paper/3244/, p. 7.

And finally, it is possible that a given indicator or measure valued by the state is not yet ready for a high-stakes framework. Given that ESSA allows states to amend their measures in future years, states might consider collecting data on measures of interest now for potential inclusion in the accountability system in the future. Student surveys, for example, are a promising way of assessing students' social-emotional skills and are being used in several districts, including California's CORE Districts, for school improvement purposes.[4] However, as discussed in greater detail below, issues may arise when using student surveys of social-emotional skills to assess students' underlying competencies in a high-stakes framework. Such measures could evolve to be part of the accountability system in a future year, or they could continue to be used solely for district and school diagnostic purposes and continuous improvement efforts.

We discuss each of the indicators and ways of operationalizing them in greater detail below.

COLLEGE AND CAREER READINESS

College and career readiness (CCR), or post-secondary readiness, is listed in ESSA as an example of a 5th indicator. A growing body of evidence suggests that CCR requires more than a strong academic foundation in math and ELA; however, articulating a concise definition for CCR remains challenging in that it is multi-faceted and context-specific. David Conley of the Educational Policy and Improvement Center (EPIC) offers four key aspects that are necessary factors for graduating students who are college and career-ready, and are therefore predictive of whether students will succeed in college and career:[5]

1. content knowledge including knowledge in the core subject areas;
2. cognitive strategies such as problem solving;
3. learning skills such as goal setting and persistence, and learning techniques including time management and study skills; and
4. transition knowledge and skills, for example, understanding which courses to take and how to apply for financial aid.

As defined by Conley, nearly every indicator that could be conceived of as a 5th indicator would fall under the CCR bucket; therefore, CCR might be more accurately viewed as a goal of the accountability system rather than as a single indicator. However, we list it here as a separate

4 The CORE Districts are a group of eight California school districts (Fresno, Garden Grove, Long Beach, Los Angeles, Oakland, Sacramento, San Francisco, and Santa Ana). They represent over one million students and 1,600 schools that are committed to working together to innovate, implement, and scale new strategies and tools that help students succeed. Six of these districts began operating under an NCLB waiver granted in August 2013 to implement a holistic system of accountability and continuous improvement focused on students' social-emotional skills and school climate/culture alongside academic outcomes. While initially intended to be used in place of California's accountability system for districts that opted in, requirements under ESSA will now take the place of the NCLB waiver. Therefore, the CORE Districts' accountability system will continue to operate, but as a no-stakes school quality improvement framework and will be known as the Data Measurement System.

5 www.epiconline.org/ccr-definition/

indicator since it is an example provided by ESSA, and states have expressed interest in thinking through how to incorporate it into their accountability system through the student success and school quality provision. Conley's definition of CCR focuses on readiness factors (i.e., leading factors) such as participation in advanced coursework or social-emotional skills, such as social awareness. CCR outcomes, or lagging factors, such as the percentage of the graduating class enrolled in a post-secondary institution or career training program could also be measured. A benefit of assessing CCR factors is that they can signal the need to intervene before the student has exited the k-12 system.[6] The benefit of assessing CCR outcomes is that high schools will be held accountable for actually producing students who are college and career ready.

For CCR to be included as the 5th indicator rather than as an overarching goal, it is important for states to carefully consider which measures will be used to operationalize it in ways that complement the data currently collected by the accountability system while also providing new data with which to hold schools and districts accountable. For example, several state leaders are considering an advanced coursework indicator (an example of a 5th indicator under ESSA) in large part because participation and success in advanced coursework has been shown to be a significant contributor to college enrollment and graduation.[7] There are two primary ways for states to operationalize the advanced coursework indicator. States can choose a measure of participation (e.g., student enrollment in AP courses or gifted/accelerated courses at the elementary school level) or a measure of performance (e.g., performance on the AP exams). Measures of participation are important because research finds that public schools serving low socioeconomic status SES communites are less likely to offer advanced coursework.[8] As such, this measure might incentivize schools to offer such courses. On the other hand, a state might consider measures that take into account actual student performance in advanced coursework in order to determine whether students will graduate with the skills ready to succeed in college.

It is important for states considering CCR as a 5th indicator to ensure they are including measures that assess both career *and* college readiness. If the measures focus on college readiness to the exclusion of career readiness, or overweight college readiness measures, technical high schools serving students who are primarily seeking career certificates may be disadvantaged by the system. Career-ready indicators that are being considered in New Hampshire, for example, include the following:

1. Student access to career pathways
2. Percentage of students completing career pathways

6 Balfanz, R., Bridgeland, J. M., Fox, J. H., DePaoli, J. L., Ingram, E. S., Maushard, M. (2014). *Building a grad nation: Progress and challenge in ending the high school dropout epidemic*. Civic Enterprises: Washington, DC. Retrieved December 30, 2016, from http://new.every1graduates.org/wp-content/uploads/2014/04/BGN_2014_Report.pdf.

7 Adelman, C. (2006). The toolbox revisited: *Paths to degree completion from high school through college*. Retrieved December 30, 2016, from the U.S. Department of Education's website, https://www2.ed.gov/rschstat/research/%20pubs/toolboxrevisit/index.html.

8 Ibid.

3. Percentage of students earning college credit that transfers to a higher education institution
4. Percentage of students earning industry recognized credentials (ICRs)
5. Percentage of students enrolled in college or secure employment within 12 and 18 months of high school graduation
6. Percentage of students participating in a work-based learning experience
7. Percentage of students pocessing work-ready employability skills

Further, it is likely that the measure used to operationalize the CCR indicator will vary by grade span (i.e., elementary vs. secondary) and school type (i.e., vocational vs. traditional). Texas, for example, collects multiple measures of CCR at each grade span (see Table 1 below). Student attendance is the only measure collected at the elementary school level, but is one of several collected at the middle school level. At the middle and high school level, measures of advanced coursework are collected. At each level, all measures feed into a final score with which to assess CCR across schools.

Table 1. Measures collected by Texas as part of their Public Reporting and Accountability System[9]

Elementary school	Middle school	High school
➢ student attendance	➢ student attendance ➢ dropout rate ➢ percentage of 7th and 8th grade students who receive instruction in preparation for high school, college, and career	➢ dropout rate ➢ graduation rate ➢ percentage of students who do at least one of the following:[10] ✓ earn at least 12 hours of postsecondary credit ✓ complete an AP course ✓ enlist in the armed forces ✓ earn an industry certification

In fact, a developmental, multiple-metric approach may be most appropriate in that research suggests that multiple factors from pre-kindergarten to high school predict whether students graduate college and career ready.[11] For example, at the middle school level, research has shown that the following factors relate to whether a student is college and career ready: the student has less than 20 percent absenteeism in middle grades; the student attends the same school through middle grades; the student passes all ELA and math courses and meets benchmarks in state exams; the student passes algebra 1 in 8th grade; the student is assigned rigorous coursework;

9 Lemke, M. (July, 2016). NGALC CCRS Measures. Presentation given at the Next Generation Accountability Learning Community. Boston, Massachusetts. Retreived from http://massinc.org/2016/08/03/ngalc-meeting-2-recap/

10 This is only a partial list of the requirements for high school students.

11 Hein, V., Smerdon, B., Sambolt, M. (November, 2013). Predictors of Postsecondary Success. Prepared for the College and Career Readiness and Success Center, American Institutes for Research. Retreived from http://www.ccrscenter.org/sites/default/files/CCRS%20Center_Predictors%20of%20Postsecondary%20Success_final_0.pdf.

and the student has strong social-emotional and decision-making skills.[12] While each measure may not be appropriate for school accountability on its own, several measures could be used as a composite index representing "on-track to readiness" at each school level. This composite could be used within each school level accountability system to incentivize best practices for increasing the probability that students will be prepared for the next level and on-track for graduating college and career ready.

Colorado, Connecticut, Florida, Texas, and Virginia are currently reporting several measures of CCR as part of their public reporting and accountability system. Given that many states already collect data related to CCR, it may require relatively little additional burden for states to create an indicator of college and career readiness using a multi-metric approach to meet the requirements of the 5th indicator. It will be important, however, for states to consider which measures to include at each grade span and whether and how to combine multiple measures within each grade span into a single metric.

SCHOOL CLIMATE

ESSA also lists school climate as an example of the 5th indicator. Measuring school climate and culture requires assessing the quality and environment of a school with regard to the level of safety, trust, and connectedness experienced by students and teachers. Research indicates that schools where students feel socially, emotionally, and physically safe are more likely to foster student engagement and academic learning. Conversely, schools where students feel unsafe or unsupported are more likely to have higher rates of chronic absenteeism and more frequent behavioral problems.[13] Thus, school climate is an important factor to consider for states seeking to improve either school quality or student success.

As listed in Table 2 below, two types of measures are typically used to assess school climate — administrative data and survey data. Both have their strengths and limitations. Administrative data, such as suspension rates, while arguably more objective than student surveys, will likely not be able to capture the full construct of school climate. Further, as discussed above, it is possible for educators to game the data (i.e., reducing suspensions) without addressing the underlying issues.

Surveys, on the other hand, can provide a more complete picture of a school's climate by asking students about the different facets that make up the school culture (i.e., feelings of belonging, feelings of safety). And, many of the climate surveys that have been developed and validated, such as the U.S. Department of Education's School Climate Surveys (EDSCL), are free and available for public use. However, since surveys are based on individual perceptions, they are accompanied by a number of problematic measurement issues such as reference bias and social desirability bias. Moreover, guaranteeing adequate survey response rates to ensure generalizability of the results

12 Lemke, M. (July, 2016). NGALC CCRS Measures. Presentation given at the Next Generation Accountability Learning Community. Boston, Massachusetts. Retreived from http://massinc.org/2016/08/03/ngalc-meeting-2-recap/

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

could be problematic. For example, students who are routinely absent may be less likely to take the survey, and since chronic absenteeism is likely related to student perceptions of the school, the survey results may be biased upwards, leading to false inferences of school climate. Finally, as with any indicator based on self-reported data, a primary consideration is to ensure that school climate surveys are not being gamed when used for high-stakes decisions. In fact, developers of some surveys on school culture and climate, like the Comprehensive School Climate Inventory (CSCI) and the EDSCLS, do not recommend using the surveys for high-stakes decisions in large part because it is unclear whether they provide reliable across-school comparisons.

Table 2. Measures of School Climate[14]

Measure	Scale/Metric	Relationship between measure and indicator of school climate	Examples of states/districts using measure
Administrative data	Suspension/Expulsion Rate	Suspension/expulsion rates, particularly disaggregated by subgroup, can indicate issues of school equity.	CORE Districts
	Chronic Absenteeism	Chronic absenteeism can provide an indirect measure of how worthwhile students think school is, whether or not they feel safe, and whether or not they have positive social interactions.	CT, NH, CORE Districts
Surveys (student, teacher and parent)	Interpersonal Relationships	A measure of the quality of relationships can indicate how supportive teachers are, feelings of social inclusion, and feelings of being valued.	CORE Districts, IL (5Essentials), GA (My Student Survey)
	Safety	A safety measure can indicate how safe (physically and emotionally) students feel attending school.	CORE Districts, IL (5Essentials), GA (My Student Survey)
	School Community Engagement	A school-community engagement measure can indicate how supportive and respectful the larger learning climate is.	CORE Districts, IL (5Essentials), GA (My Student Survey)

Despite potential concerns with survey data, research has shown that survey results can be highly reliable and predictive of other outcomes.[15] States such as Connecticut and New Hampshire are already administering surveys on school climate for public reporting, and other states such as California and Tennessee are considering including school culture surveys as a measure for their 5th indicator. Further, several districts currently collect school climate data for accountability and reporting purposes (e.g., Austin Independent School District, Texas; Washoe County School District, Nevada; the CORE Districts, California).

14 Adapted from Lipman, A., Bundy, A., Forman, B., Reyna, R., and Baehr, K. (2016 Jul). Unpublished Briefing Memo #2: *Accountability in Three Areas of School Quality/Student Success: College & Career Readiness, Social and Emotional Learning, and School Climate*. Boston, MA: Next Generation Accountability Learning Community.

15 See http://k12education.gatesfoundation.org/wp-content/uploads/2015/12/Asking_Students_Practitioner_Brief.pdf.

The CORE Districts offer one example of efforts to include a school culture and climate indicator in a school improvement framework. To mitigate issues related to bias and corruptibility, the CORE Districts use a multi-metric approach to measuring school climate that includes student, teacher, and parent surveys with chronic absenteeism, and suspension and expulsion rates. Their school culture and climate items, which draw heavily from the California Healthy Kids Survey developed by WestEd, were field tested in 2014-15 with nearly half a million students. Given the strong validity and reliability evidence collected from the field test, the CORE Districts moved forward with including the surveys in their school improvement system with the recognition that validation efforts would continue in order to ensure that gaming did not occur as the system became fully operational.

STUDENT SOCIAL-EMOTIONAL MINDSETS AND SKILLS

In recent years, there has been a strong emphasis on having schools focus on student social-emotional skills in addition to traditional academic content. Social-emotional skills include constructs such as self-management and social competence, and mindsets like self-efficacy and growth. Research has shown that social-emotional competencies predict important academic outcomes, including college persistence, as well as career and life outcomes.[16] Further, research has shown that teachers who are effective at raising test scores are not necessarily the same teachers able to improve students' social-emotional skills, but that high school graduation rates are higher in schools with more teachers who are adept at raising both.[17] And, research has found that improving social-emotional skills can temper the effect of poverty on academic achievement.[18] Therefore, measuring students' social-emotional skills through ESSA would provide an important way for states to signal to schools and districts that these skills are important to develop in addition to academic skills and content knowledge. [19]

There are two types of measures that can be used to assess students' social-emotional skills — those based on ratings of student behavior (i.e., self-reported student surveys or teacher surveys of student competencies) and those based on direct assessment of the skill through performance tasks or educator observations. The former — student and teacher surveys — are more widely used. For example, several districts throughout the U.S. are administering surveys of

16 For more information on the longitudinal research base showing the impact of social-emotional competencies on students' outcomes, see Gabrieli, C, Ansel, D., and Krachman, S. B. (2015). *Ready to be counted: The research case for education policy action on non-cognitive skills*. Retrieved December 30, 2016, from http://www.transformingeducation.org/s/ReadytoBeCounted_Release.pdf.

17 Jackson, K. (2016). *What do test scores miss? The importance of teacher effects on non-test score outcomes.* (Working paper, 22226). National Bureau of Economic Research: Cambridge, MA: Kraft, M., and Grace, S. (2016). *Teaching for tomorrow's economy? Teacher effects on complex cognitive skills and social-emotional competencies.* (Working paper). Retrieved December 30, 2016, from http://scholar.harvard.edu/files/mkraft/files/teaching_for_tomorrows_economy_-_final_public.pdf.

18 Claro, S., Paunesku, D., and Dweck, C. S. Growth mindset tempers the effect of poverty on academic achievement. Proceedings of the National Academy of Sciences, *113*(31): 8664-8668.

19 Gabrieli, C, Ansel, D., and Krachman, S. B. (2015). *Ready to be counted: The research case for education policy action on non-cognitive skills*. Retrieved December 30, 2016, from http://www.transformingeducation.org/s/ReadytoBeCounted_Release.pdf.

students' social-emotional skills including include Washoe County School System (Nevada), Austin Independent School District (Texas), and Anchorage School District (Alaska).[20] Further, the National Assessment of Education Progress (NAEP), which is administered to a nationally representative sample of students each year, is including measures of social-emotional skills (such as perseverance and desire for learning) in the 2016-17 school year.

Self-reports do have their limitations, particularly if used for accountability. As discussed in a paper by Angela Duckworth and David Yeager, there are several issues that can affect how students rate their behavior, which can introduce bias into the scores.[21] For example, some students may rate themselves higher due to social desirability bias, or the tendency for survey responses to be influenced by *social pressures*.[22] Another issue is that of stereotype threat, whereby survey responses may be influenced by respondents' *perception* of how people in their group (e.g., racial, ethnic, or socio-economic class) *are believed by others to perform* in that competency area.[23] A third issue is that of reference bias, whereby one's responses are influenced by one's individual frame of reference. This can lead to a situation in which students with the same underlying skill level interpret the scale differently and therefore rate themselves differently, potentially making it difficult to compare average scores across schools.[24] And finally, a consideration with any type of self-report is that it can easily be gamed to indicate that students have attained a higher level of a given competency than they actually have.

Performance tasks and observations of direct behavior, including game-based tasks, are another way of measuring students' social-emotional skills. A common example of a performance task of students' self-control is the widely cited "marshmallow test" administered by psychologist Walter Mischel to children of Stanford faculty and graduate students in the 1960s. The experiment involved offering a child a small treat to eat immediately (e.g., one marshmallow) or a larger treat to eat (e.g., two marshmallows) if the child could wait until the researcher returned to the room (up to 20 minutes later). Whether and how long the child was able to wait to eat the larger treat has been shown to be strongly associated with desirable later-life outcomes, such as higher SAT scores and lower body mass index (BMI).[25] A more recent and more scalable example comes

20 These are three of eight urban districts that have committed to implementing comprehensive social-emotional learning practices for students in elementary, middle, and high schools through an Institute of Education Sciences grant coordinated by the Collaborative for Academic, Social, and Emotional Learning and with assistance from the American Institutes for Research. See http://www.edweek.org/ew/articles/2015/06/10/urban-districts-embrace-social-emotional-learning.html.

21 Duckworth, A., and Yeager, D. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4): 237-251.

22 Podsakoff, P. M., MacKenzie, S. and Lee, J-Y. L. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879.

23 Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(4): 4-28.

24 Dobbie, W., & Fryer, R. G., Jr. (2013). *The medium-term impacts of high-achieving charter schools on non-test score outcomes*. NBER Working Paper Series. Cambridge, MA: National Bureau of Economic Research; Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, *82*(6), 903–918; Goldman, S. (2006). Self-discipline predicts academic performance among low-achieving adolescents. *Res: A Journal of Undergraduate Research*, *2*(1), 84–97.

25 Mischel, W. (2014). *The marshmallow test: Mastering self-control*. New York, NY: Little, Brown and Company.

from the Programme for International Student Assessment (PISA), an assessment administered internationally through the Organisation for Economic Co-operation and Development (OECD). In 2015, PISA piloted a computer-based task, which in addition to measuring academic skills is intended to measure social-emotional skills such as collaborative problem solving.

Given that performance tasks do not rely on the self-report of individuals, they are less susceptible to the forms of bias mentioned above. However, performance tasks also have limitations that may influence their usefulness in an accountability system. In particular, it is difficult to design a task to ensure that it is only measuring the underlying construct intended to be assessed.[26] For example, it is possible that the marshmallow test is actually measuring students' trust in adults (i.e., trust in whether the researcher will actually give the student the larger treat when he/she returns, or whether he/she will return at all) or students' willingness to please the researcher.[27] Both factors can lead to construct irrelevance or task impurity, which makes it difficult to determine if a student's score on the performance task is actually representative of their true underlying ability. Moreover, while a task may provide an objective measure of a student's ability, the scoring of the behavior (by researchers or educators) may be subjective. Finally, performance tasks do not always provide insight into typical behavior since a task that optimizes students' motivation to perform well may not reflect behavior in day-to-day situations. At this point, we know far less about how performance tasks would play out as a measurement of students' social emotional skills because they have not yet been piloted at scale.

The CORE Districts offer one example of how social-emotional skills through student surveys can be used in a school improvement framework. The CORE Districts chose to measure four social-emotional competencies through student surveys in their school improvement system: growth mindset, self-efficacy, self-management, and social awareness. The scales were developed by leading researchers and are free to use for educational purposes through Transforming Education's report Measuring MESH. They can be administered inexpensively at scale.[28] Preliminary research on the CORE Districts' results suggests that measuring students' social-emotional skills through student self-reports can provide valid and reliable data on student skills that is not captured by either school culture surveys or academic outcomes. [29] Evidence of this comes from the largest rollout of student surveys of social-emotional skills to date from the CORE Districts 2014-15 field test of their new accountability system, whereby data from nearly half a million students was collected and assessed. In fact, results suggest that the self-reports of students' social-emotional skills would meet ESSA requirements of validity, reliability, comparability, and meaningful differentiation.[30] However, because this data was collected during the no-stakes field test, it is

26 Duckworth, A., and Yeager, D. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4): 237-251.

27 Kidd, C., Palmeri, H., and Aslin, R. N. (2016). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition, 126*(1): 109-114.

28 http://www.transformingeducation.org/measuringmesh/

29 Transforming Education. (2016, Apr). *Measuring MESH: Student and teacher surveys curated for the CORE districts*. Author: Boston, MA.

30 West, M. (2017, March). Should non-cognitive skills be included in school accountability systems? Preliminary evidence from California's CORE Districts. *Evidence Speaks Reports, 1*(13). Retrieved December 30, 2016, from https://www.brookings.edu/wp-content/uploads/2016/07/EvidenceSpeaksWest031716.pdf.

unclear how the results might change under a different kind of system, for example, one that uses punitive consequences in a high-stakes fashion.

Measures of students' social-emotional skills may not be ready for high-stakes accountability, yet there is a growing consensus that these skills are incredibly important in their own right and for academic outcomes. One option is to include results from measures of social-emotional skills for public reporting in order to empower educators to have more informed conversations about students' well-being. This will also enable states to learn more about which measures are the most valid and reliable. Education leaders can then continue to gather and examine several years' worth of data before deciding whether to embed a given measure into a high-stakes accountability system. In the meantime, states and districts will have the ability to use data collected as part of needs assessments to target specific supports and interventions for struggling schools.

PATHWAYS TO INNOVATION: MULTIPLE INDICATORS, MULTIPLE MEASURES, AND ADAPTIVE MEASURES

What is unique about ESSA is that it offers states a myriad of ways to expand their definition of student success and school quality through the 5th indicator. In this section, we discuss three potential pathways being explored in practice by the states contributing to this brief. These three pathways include multiple indicators (Connecticut); multiple measures for a single indicator (Tennessee); and "adaptive measures" such as including inputs of an indicator that vary from school to school, district to district, and even year to year (West Virginia). To show how these pathways play out in practice, we include case studies written directly by the identified state. Each state has been thoughtful and deliberate in their approach to including new indicators that can satisfy ESSA's 5th indicator requirement and have begun to conceptualize accountability systems that go beyond mere compliance so as to have the greatest effect on student outcomes. All three states have expressed the need to both continue to evaluate and refine their system through the coming years to ensure the system allows for a fair and accurate assessment of school performance, and to remain open to changes of their existing indicators or measures based on new research that emerges.

MULTIPLE INDICATORS

Instead of focusing on only one additional indicator of school quality or student success, states may choose to include several indicators. If a state does choose more than one indicator of school quality or student success, each indicator can be thought of as a separate and distinct factor that is important for student success; however, the indicators should collectively be viewed as being in service to a particular goal or multiple goals of the accountability system. Marion and Lyons, from the Center for Assessment, offer a conceptualization of indicators of school quality or student success by the information they provide to the system: pre-cursor indicators, school quality indicators, and school success indicators. Precursor variables are causally linked to predicting

academic achievement.[31] On the other hand, school quality and student success indicators are value propositions about what characterizes a high-quality school or successful student.

Table 1: Categorization and examples of Indicators for School Quality and Student Success[32]

Pre-cursor Examples	School Quality Examples	Student Success Examples
➢ Algebra readiness ➢ Credits earned by end of 9th grade ➢ Enrollment in advanced/ accelerated coursework	➢ Student/community engagement ➢ School climate ➢ Extra-curricular participation ➢ Arts participation ➢ Educator quality (qualifications, experience, effectiveness) ➢ Quality of local assessments or assessment practices ➢ Engagement in professional capacity building	➢ Social-emotional skills ➢ Physical fitness ➢ Earning a career/technology certificate ➢ Earning college credit ➢ Persistence in post-secondary education ➢ Entering STEM field

Connecticut is one state that has proposed several 5th indicators related to each of the three categories identified by Marion and Lyons. For example, among the dimension of pre-curser indicators, Connecticut includes the percentage of students passing advanced coursework examinations. Along the dimension of school quality, Connecticut includes the percentage of students with access to the arts. And, along the dimension of student success, Connecticut includes the physical fitness of students. In total, Connecticut has 12 indicators, each contributing to a 1250-point index. All indicators are in service to the following goals:

- Provide a more complete picture of a school or district
- Guard against narrowing of the curriculum to the tested subjects
- Expand ownership of accountability to more staff
- Allow schools to demonstrate progress on "outcome precursors"[33]

Box 1, below, provides greater detail on Connecticut's multiple indicator system and discusses the state's rationale for including multiple indicators. The state has further indicated its openness to considering new indicators to include in the system in future years depending on stakeholder interest and results from pilot tests.

31 Marion, S. and Lyons, S. (2016). In Search of Unicorns: Conceptualizing and Validating the "Fifth Indicator" in ESSA Accountability Systems. Dover, NH: National Center for the Improvement of Educational Assessment. Retreived from http://www.nciea.org/publication_PDFs/Marion%20Lyons_ESSA%20Accountability_5th%20Indicator_111416.pdf

32 Ibid.

33 Connecticut State Department of Education. Using accountability results to guide improvement. Retreived from http://www.sde.ct.gov/sde/lib/sde/pdf/evalresearch/using_accountability_results_to_guide_improvement_20160228.pdf.

Box 1: Connecticut's Approach to Multiple Indicators under ESSA

The Connecticut State Department of Education's (CSDE) espoused theory of action for education is summed up in the following statement from its comprehensive plan:

"If Connecticut is successful in developing and supporting the structures and conditions that create a culture of high expectations for all students, developing great teachers and leaders, and supporting great schools, then the results will be threefold: we will improve outcomes for all students; close achievement gaps; and prepare well-rounded, engaged students who graduate ready to succeed in college, career, and civic life."

To create an accountability system that aligns with this theory, the CSDE actively sought feedback from district and school leaders, Connecticut educators, state and national experts, state staff, and many others for over a two-year period. Finally, in March 2016, Connecticut launched the Next Generation Accountability System that more fully operationalized the theory of action articulated in its comprehensive plan.

This system incorporates several indicators to provide a more holistic, multifactor perspective of district and school performance. In addition to test scores and graduation rates, it focuses on chronic absenteeism, access to college/career readiness coursework, achievement on college readiness exams, percentage of 9th graders earning at least five credits, a 6-year graduation rate for our most vulnerable students, access to the arts, physical fitness, and postsecondary entrance. With respect to the state summative assessment indicators, the system utilizes performance and growth indices based on scale scores to permanently eliminate the focus on students on the proficiency "bubble." The system also weights subgroup performance separately on achievement, growth, chronic absenteeism, and graduation such that a school/district cannot be high performing overall without a high performing subgroup comprised of an unduplicated count of English learners, students with disabilities, and students receiving free and reduced-price lunch. The system accomplishes all of this without collecting any new data from districts.

Connecticut completed the design of its accountability system well before the passage of ESSA. The system happens to align well with ESSA's requirements for indicators and weights. Connecticut does not view the 5th indicator as an afterthought or a compliance exercise; on the contrary, the many indicators in Connecticut's model that could satisfy the federal requirement derive meaning from their alignment to Connecticut's comprehensive plan. This alignment is vital because the indicators don't stand in isolation — they connect to state priorities and strategic actions. For example, the Connecticut General Assembly's Committee on Children is promoting a systemic response that involves state and local education and community partners to combat chronic absenteeism and encourage student attendance. The inclusion of chronic absenteeism as an indicator is fostering a sense of shared accountability in these efforts.

The broadening of indicators is expanding ownership of accountability in a school or district to many more staff — no longer is accountability limited to the work of ELA and math teachers in the tested grades. Though the accountability system is still in its infancy, there are signs that senior

leaders at the local level are taking a closer look at student attendance, revisiting advanced course offerings and placement practices, renewing efforts to re-engage students at risk of dropping out, expanding arts offerings, and reflecting on the physical fitness and overall health of their students.

Transforming accountability from a compliance-driven, rank and sort exercise into one that is viewed as providing useful information for decision making is unending work. To truly engender trust from practitioners, this belief is affirmed in every communication and presentation about the accountability system. In fact, when Connecticut released the first results last March, in addition to scores, lists, indexes, and percentages, it also released a comprehensive guide for using accountability results to inform local improvement efforts; this guide offers links to resources, research, and evidence-based strategies for each indicator in the system and connects practitioners to CSDE experts who can help them improve in each domain.[34]

Though Connecticut is well positioned to implement ESSA's accountability requirements, they remain open to learning from their experiences and the results, and are committed to continuous improvement. Specifically, when new ideas are proposed, CSDE will evaluate them as part of their multi-measure system, and if appropriate, incorporate them into the system in a future year. However, the incorporation of new measures will require thoughtful deliberation to balance continuous improvement with having a consistent set of indicators for measuring trends in performance over time.

MULTIPLE MEASURES

In addition to including multiple indicators that assess different aspects of the state's theory of action, states might consider multiple measures for a multi-metric indicator. The use of multiple measures is advised in *The Standards for Educational and Psychological Testing*, which states, "In educational settings, a decision or characterization that will have major impact on a student should not be made on the basis of a single test score. Other relevant information should be taken into account if it will enhance the overall validity of the decision."[35] Including multiple measures can enhance construct validity by providing additional data points with which to assess an underlying construct, while also reducing the likelihood that any decisions about a school will be driven by factors outside of the school's control.[36] Further, it can reduce the risk of corruption since each measure would likely only contribute a small amount to the overall score on the indicator.[37] On the other hand, including multiple measures may lead to less actionable feedback

34 See http://raguide.org/turn-the-curve-exercise-results-based-accountability/.

35 American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA/APA/NCME]. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association. Specifically Standard 13.7.

36 http://www.ascd.org/publications/educational-leadership/nov09/vol67/num03/The-Many-Meanings-of-%C2%A3Multiple-Measures%C2%A3.aspx

37 Marion, S. (2016, February). *Considerations for state leaders in the design of school accountability systems under the Every Student Succeeds Act.* National Center for the Improvement of Educational Assessment: WestEd and CRESST. Retrieved December 30, 2016, from http://www.nciea.org/publication_PDFs/ESSA%20Accountability%20Design%20Considerations_021916.pdf.

and spread resources too thinly. Further, given that the additional indicator must count for significantly less than academic indicators, including multiple measures runs the risk of watering down individual measures.

Tennessee is considering one of two indicators to comply with the 5th indicator requirement under ESSA — either "college and career readiness" under the student success category or "equity and opportunity" under the school quality category. The state has begun the process of considering which measures would comprise each indicator as part of a multi-metric approach. With the indicator of equity and opportunity, for example, measures might include chronic absenteeism, discipline data (including suspension rates), and access to effective teachers. These measures would roll up into a single score, intended to convey different aspects about the same underlying construct of student equity and opportunity. As noted below, there are several benefits of including multiple measures. However, there are also challenges, including questions of how to combine the measures to inform a single indicator (i.e., issues of weighting) and concerns with ensuring that any final categorization provides actionable information to educators. These challenges are not inconsequential and they will require careful consideration prior to rolling out the new indicator.

The Tennessee Department of Education describes key considerations for their choice of indicator and associated measures, highlighting the benefits and limitations of a multi-measure approach in Box 2 below.

Box 2: Tennessee's Consideration of a Multiple-Measures Index

Our theory of action assumes that schools and districts will be empowered in using their data to shape practice and target intervention. As such, one of Tennessee's chief concerns with including multiple measures as part of such a composite indicator is the feedback that schools and districts receive from such an evaluation. Therefore, we have been looking at two main thematic buckets when considering an additional indicator index to support schools and districts in creating actionable steps from accountability evaluations.

College and Career Readiness Indicator

A postsecondary readiness indicator could include some or many of the following individual measures:

- The number and percent of students earning industry certifications
- The number and percent of students graduating with at least two university credits (as earned through dual enrollment, scoring a 3 or better on an AP exam, etc.)
- The number and percent of students completing career and technical education course progressions
- The number and percent of students still enrolled in postsecondary two years after graduation (using the denominator of either all college-going students or all students in the graduating cohort)

In Tennessee, there are a handful of challenges associated with constructing such an index:

- Industry certifications are tracked at the state and national levels, making it difficult to associate such certificates with individual student records.
- There are inherent differences in the early postsecondary opportunities (EPSOs) offered in different districts, in addition to the fact that credit attainment rates look different across different categories of EPSOs (e.g., AP exams vs. local dual credit).
- Many students take a sampling of CTE courses rather than progressing through a single track, which raises questions as to how to account for such trajectories.
- Including postsecondary outcomes after high school graduation raises questions about the utility of such feedback in creating actionable steps for schools and districts. In other words, introducing lagging measures of postsecondary attainment hinders schools' and districts' ability to intervene to correct problematic situations, as they do not currently serve the students for whom they would be held accountable. It may consequently be difficult to elucidate meaningful action steps from such lagged data.

Equity and Opportunity Indicator

An equity and opportunity indicator could include some or many of the following individual measures *with an emphasis on disparity and disproportionality between subgroups*:

- Chronic absenteeism
- Discipline
- Access to effective teachers

As with our investigations into a multi-measure college and career readiness indicator, there are some potential weaknesses of a multi-measure equity/opportunity indicator as well:

- Some districts have stated that chronic absenteeism is a data point that exceeds districts' locus of control.
- Discipline data vary widely within and between districts; suspension rates alone do not address, say, restorative justice practices or underreporting.
- Access to effective teachers, in the way the analysis has proceeded in Tennessee, requires minimum counts of students achieving in the top and bottom quartiles, as well as minimum counts of teachers who are highly effective and not, as measured by our value-added measure. This has incited some pushback from stakeholders.

The benefits and challenges of a multi-measure approach with either indicator will be considered in greater detail prior to the full rollout of Tennessee's system. A primary consideration is the weighting of each indicator in a final school ranking. One view argues that because the 5th indicator accounts for less weight, using multiple measures may effectively "water down" the importance of each individual measure, thereby creating barriers or disincentives for school and district growth in

these areas. This may, however, act as a positive outcome if there is little consensus around which measures belong in a high-stakes evaluation framework. Including multiple measures can produce a palliative effect for those measures that cause concern among certain groups.

An opposing viewpoint argues that the seeming lack of consensus in the research base, due to its nascent nature, supports the inclusion of multiple measures given that a suboptimal outcome from a single measure stands to have a larger chance of affecting a school or district's overall evaluation. That this indicator, whether it contains multiple measures or a single one, still must account for much less than the academic indicators in a school's performance evaluation, is likely some comfort to pessimists looking to do as little harm as possible and is also perhaps little reward for optimists seeking incentives for reform. Certainly, while a focus on worst-case scenarios seems anathema to the strategic approach toward continuous improvement that underlies the development of Tennessee's accountability framework, this is still a perspective worth noting as the pitfalls of additional indicators loom large in considerations around revamping accountability systems.

ADAPTIVE MEASURES

A third option states might consider is that of "adaptive measures," whereby the measures being used for a particular indicator might differ from school to school, district to district, and even from year to year, in order to ensure that the chosen set of measures provides the greatest degree of construct representation. A particular consideration under the "adaptive measures" framework is to meet ESSA's requirement of comparability. Comparability of scores ensures that schools can be meaningfully compared on an indicator, even if the way the indicator is operationalized is different across schools. Since the final regulations do not require the same measure across schools, regardless of grade span, the adaptive measure approach seems feasible under the law.

West Virginia's conceptualization of the school quality or student success indicator, *at-risk subgroup reduction*, provides an illustrative example of an adaptive measures approach (see Box 3). Recognizing that risk factors of students dropping out can differ across schools and districts depending on the student population, West Virginia utilizes an algorithm that takes into account historical longitudinal data and profiles of students who have dropped out specific to each school and school district. Statistical models are used to calculate highly accurate risk predictions based on the past 12 months of data. This predictive analytics approach has yielded a 90 percent accuracy rate of identification of potential dropouts in the elementary grades and up to 98 percent accuracy in the high school grades, which is substantially more accurate than traditional checklist/threshold approaches.

Based on this algorithm, the measures used to predict whether a student is at risk for dropping out might vary across schools. Further, as schools and districts in West Virginia address those risk factors, new factors might emerge with greater predictive power. It is feasible, therefore, that the measures used to identify students of being at-risk will change from year to year, as the algorithm identifies new factors over time. In this way, the use of the adaptive measures, as conceptualized under West Virginia's 5th indicator, can actually be seen as *improving* the validity and comparability of the indicator.

Box 3: Incorporating Drop-out Prevention into School Accountability: West Virginia's Early Warning System

In recent years, increasing graduation rates has become a central focus for the public education system in West Virginia. As such, identifying students at risk for dropping out of school and implementing interventions at earlier ages have become crucial to helping more students graduate. Recognizing the limited time and resources school personnel may have to independently compile, analyze, and draw conclusions about the drop-out potential for students possessing certain risk factors, the West Virginia Department of Education (WVDE) partnered with Bright Bytes, Inc., to develop an early warning system to individualize dropout prediction and prevention. Drawing on decades of longitudinal data, an advanced, research-based algorithm was deployed to determine the factors that drive dropouts that are unique to each school. Multiple data points spanning the domains of academics, attendance, behavior, and demographic characteristics are used to predict the drop-out risk for each student in grades 1 through 12. The system was fully implemented statewide during the 2015-2016 school year.

By the time ESSA was passed in December of 2015, stakeholders in West Virginia already were engaged in discussions around redesigning the state accountability system to include a wider spectrum of measures. With its provisions to include not only academic indicators but indicators of school quality or student success, ESSA inspired stakeholders to think more broadly about potential measures. The state's focus on increasing graduation rates, coupled with ESSA's requirement for including graduation rates into high school accountability, spurred conversations about rewarding the contributions that elementary and middle schools make toward students' potential for successfully completing school (i.e., an elementary/middle school measure corresponding to high school graduation). As a result, a recommendation was put forth to the West Virginia Board of Education to incorporate dropout risk reduction into the accountability system. Ultimately the state board adopted the recommendation for all school programmatic levels, but weighting it more heavily at the elementary and middle school levels.

The *at-risk subgroup reduction* accountability indicator demonstrates the extent to which schools reduce dropout potential among students identified at the beginning of the school year as being subject to moderate to high levels of risk. With the risk predictions generated by the West Virginia Early Warning System (EWS), educators can intervene early and individualize evidence-based interventions corresponding to student's particular risk factors. Student progress can be tracked in an ongoing, real-time basis in the EWS and through ZoomWVe for Educators, which provides school personnel in the state with secure access to a broad spectrum of student data for use to improve pre-kindergarten through grade 12 students' learning. For purposes of school accountability, a two-stage process is used whereby schools are rewarded for the proportion of students with no or low dropout risk, and for the extent to which they can reduce the dropout potential of at-risk students over the course of the school year.

MEASUREMENT CONSIDERATIONS

Redefining the meaning of accountability to extend beyond NCLB's requirements, and choosing non-academic indicators that align with the state's theory of action are important steps toward creating a more holistic vision of student success. An additional step is identifying which measures will be used to operationalize the 5th indicator and how those measures will be rolled up into a metric reported in the accountability system. Here, we focus on technical aspects of the measures for states to consider as they seek to operationalize their chosen 5th indicator.

IDENTIFYING WHERE TO FOCUS MEASUREMENT ACTIVITY

To determine where to focus measurement activity, states might think about the choice of measure along two dimensions — level of inference and level of construct representation. Level of inference indicates how far removed the measure is from readily observable data. Level of construct representation refers to the extent to which the measure assesses the breadth of the underlying construct. Often, measures that have higher levels of construct representation, such as student survey scores, contain higher levels of inference, and therefore require a "heavier burden of evidence to support their validity."[38] Measures with lower levels of inference, such as counts based on administrative data (i.e., absenteeism rate), may be less susceptible to corruptibility since they can be verified with objective data. Further, they often place a lower level of data burden on schools since they may already collect this information. However, the drawback of such measures is they may be more distally related to the chosen indicator.

Take, for example, the indicator of student social-emotional skills. States, using the CORE Districts as their model, may choose to measure social-emotional skills, such as self-management, through student surveys. A self-management survey scale can contain a variety of items and therefore has the potential to offer a high degree of construct representation. However, self-reports typically require a higher level of inference, since the student's self-management skills are not being directly observed, but rather a series of student responses to items are used to estimate the student's skill level in this area. States might instead choose to measure self-management as the percentage of times a student turns in homework on time. While this is a more objective measure than student self-reports and requires a lower level of inference, it does not cover the full breadth of self-management since self-management is demonstrated by other behaviors in addition to turning in homework on time. Therefore, the homework completion metric, while perhaps a useful proxy, is likely more distally related to the underling construct of self-management. Incentivizing behaviors that are only peripherally related (e.g., homework completion) may distract focus from moving the needle on the construct itself (e.g., self-management); moreover, the relationship between such distal indicators and longer term outcomes (e.g., high school graduation) may be weaker.

38 Marion, S. and Lyons, S. (2016). In Search of Unicorns: Conceptualizing and Validating the "Fifth Indicator" in ESSA Accountability Systems. Dover, NH: National Center for the Improvement of Educational Assessment. Retreived from http://www.nciea.org/publication_PDFs/Marion%20Lyons_ESSA%20Accountability_5th%20Indicator_111416.pdf

ALIGNING VALIDATION EFFORTS TO A THEORY OF ACTION

An additional consideration with the inclusion of any measure in an accountability system is whether the measure has sufficient evidence of validity and reliability. Validating an assessment for its intended use requires ensuring that the scores accurately capture performance on the construct the assessment is intended to measure. Many different types of evidence can and should be brought to bear to show that assessment scores can be used to support their intended interpretation and uses within an accountability system. Content-related validity evidence (i.e., evidence showing that the assessment adequately covers all sub-components of the construct it's purporting to measure) and criterion-referenced validity evidence (i.e., correlational evidence showing that scores on the measure align with other outcomes that the construct is hypothesized to be related to) are two types of evidence. Additional sources of validity evidence include evidence based on response processes (e.g., results from student cognitive interviews); evidence based on internal structure, including reliability; and evidence related to the consequences associated with test use.

When investigating predictive validity, any examination of how the 5th indicator correlates with other student outcomes, including academic and behavioral outcomes, should be interpretted within the context of whether and how much the two indicators are expected to align. As such, part of validating a measure is using a theory of action to map out the ways in which the constructs are believed to align to other outcomes. One might not expect the 5th indicator to correlate strongly with academic outcomes if it is intended to measure student success or school quality along dimensions that are different from academic achievement. As Marion and Lyons note, "If an indicator represents something distinct from traditional test-based as academic achievement, then we would not necessarily expect a strong relationship between assessment results and available performance on this indicator".[39] In fact, perfect, or even strong, correlations between the 5th indicator and academic test scores, for example, can suggest that the additional indicator is unnecessary since it is providing little additional information to the accountability system that is not already captured by test scores.

To that end, there are a variety of indicators of school quality that might not directly correlate with academic performance, but may be important for state goals nonetheless. These include process-based indicators (e.g., teacher hiring practices), indicators that incentivize best practices (e.g., engagement in professional capacity building), and precursors to student success (e.g., school climate). An example of an indicator that incentivizes best practices is that of equitable access to AP courses. While such an indicator may not immediately drive student academic outcomes, it may be important to include in a state accountability system that is focused on ensuring equity of opportunity and increasing access to credit-bearing college-level coursework. Further, including this indicator in an accountability system can help ensure that all high schools work to provide rigorous courses. The theory of action may articulate an expectation that with increased equity in access to higher-level coursework, other indicators of college and career readiness such as college enrollment would improve. A validation of this theory of action would

39 Ibid, p. 20.

begin with collecting data on AP course-taking and college enrollment pre-and post-inclusion of this indicator in the accountability system.

After developing a theory of action delineating the ways in which the state expects the measure to correlate with other academic outcomes, an important next step in the validation process is to pilot the chosen measure in order to collect validity and reliability evidence, and identify any unintended consequences that may arise (e.g., manipulation of scores). Connecticut, for example, is considering additional indicators to potentially include in future years, such as school climate, beyond the 12 indicators they've currently identified as part of their accountability system. However, Connecticut has committed to piloting all additional measures before implementing them to ensure that all measures have been validated.[40]

DETERMINING HOW TO CALCULATE THE METRIC

With any chosen measure, states will need to determine how to calculate and report the metric. With measures based on administrative data, decisions will need to be made about which group of students are included in the count (i.e., the denominator). Decisions regarding the group of students for whom the metric is calculated can affect how a school is perceived to have performed and can alter the incentives facing the school. For example, in calculating a college matriculation rate, the denominator could consist of all 9th graders, all seniors, or all seniors who graduate. If a school has a large population of students who drop out, but a high percentage of graduating seniors who attend college, that school will look particularly poor if the denominator is based on the total number of 9th graders. But, that school will look particularly positive if it is based on the total number of students who graduate high school. Choosing a denominator based on the total number of 9th graders essentially double-counts student dropouts against a high school's evaluation, while choosing a denominator based on the total number of students who graduate may incentivize schools to encourage low-performing students to drop out in order to increase their score on the indicator. With indicators measured through surveys, decisions will need to be made regarding how to calculate scores. For example, the CORE Districts have chosen to calculate results from student surveys of social-emotional skills and school culture-climate as the percentage of positive responses (i.e., the percent of students who indicate a 4 or 5 on a 5-point Likert scale), rather than as an average score.

Since decisions about how to calculate a measure often do not involve a simple linear transformation of the scale, they can alter inferences of school performance and thus can change the incentives school face. This issue was recently discussed in an open letter to the U.S. Department of Education that sought to provide guidance on the metric used for academic test scores. Morgan Polikoff (2016) argues that test scores averages, instead of percent of proficient students, provide better incentives for schools for the following reasons:

40 Connecticut State Department of Education. (March, 2016). Connecticut's Next Generation Accountability System 2014-15 Results. Retreived from http://www.sde.ct.gov/sde/lib/sde/pdf/evalresearch/next_generation_accountability_system_march_2016_sbe_presentation.pdf

1. Percent proficiency incentivizes schools to only focus on students around the cut-off of proficiency
2. Percent proficiency encourages teachers to raise student performance to a minimum level of proficiency rather than to advanced levels
3. Percent proficiency doesn't reward schools that make large improvements in performance below the proficiency cut-off[41]

COMBINING MULTIPLE INDICATORS INTO A SINGLE COMPOSITE ACCOUNTABILITY SCORE

An additional measurement consideration is how to combine the various indicators for decisions about school support and public reporting. ESSA requires states to identify schools in the bottom 5 percent to receive Comprehensive Support and Improvement (CSI). The regulations require states to produce a summative score for each school, which means that states will need to consider how to weight each indicator to determine a final school-level score.

If all indicators are weighted equally, higher performance on one indicator can compensate for lower performance on another (i.e., a compensatory approach).[42] Practically speaking, this means that a school that is very low on any one indicator may not be identified as needing support. In an analysis of the CORE data by the Policy Analysis for California Education (PACE), the authors found that schools that were given a low overall ranking under an equal-weighting approach were not necessarily those with the lowest performance on a single indicator. As such, producing a summative ranking may ignore schools that need support in a certain area.[43]

Yet, as the regulations currently stand, equal weighting is not permitted, since indicators 1-4 must have greater weight than the 5th indicator(s) and since a school cannot graduate out of CSI (i.e., identification of the bottom 5 percent) based on progress on the 5th indicator(s) alone.[44] This suggests that indicators 1-4 should be weighted much more heavily than the 5th indicator.[45] Connecticut's proposed system, for example, contains 12 indicators, 6 of which would be categorized under the "5th indicator," and are each weighted at 4 percent, for a total of 24 percent

41 Polikoff, M. (2016, Jul). A letter for the U.S. Department of Education (final signatory list). Retrieved December 30, 2016, from https://morganpolikoff.com/.

42 Technically, even under an equally-weighted approach, indicators with greater variation will contribute more to the final score.

43 Hough, H., Penner, E., and Witte, J. (2016). *Identity crisis: Multiple measures and the identification of schools under ESSA*. PACE: Location unknown. Retrieved January 2, 2017, from https://edpolicyinca.org/publications/identity-crisis-multiple-measures-and-identification-schools-under-essa. The authors assert that for the 5th indicator to not move a school out of CSI, it must be weighted at 0. It is not clear from the regulations that this was the intent of the law, however.

44 At press time for this publication, the regulations had been suspended and it is unclear what, if any, regulations will replace them.

45 Hough et al., (2016) assert that for the 5th indicator to not move a school out of CSI, it must be weighted at 0. It is not clear from the regulations that this was the intent of the law, however.

in the overall index. Academic and graduation indicators make up the remainder of the index. This weighting schema guarantees that performance on academic and graduation indicators play a larger role in the final summative score than performance on the 5th indicator of school quality or student success. However, the combined total percentage of the 5th indicators should be large enough to incentivize schools to engage in best practices to promote the necessary skills that the state is trying to incentivize.

As states make decisions about whether and how to combine scores from indicators for the purposes of accountability, an important component will involve examining the extent to which the identification of schools placed in the bottom 5 percent will change when the method for combining the indicators changes. As Hough and colleagues find in their examination of different weighting schemas on the identification of schools for CSI: "Technical decisions about the identification scheme will have large impacts on which schools are identified for CSI and on whether low-performing schools [on a given indicator] are excluded from CSI Support."[46]

SYSTEM EVALUATION

The identification and implementation of the measures of school quality or student success are only part of the process. Crucial to sustainability of the system is to build in a plan for evaluation in order to continue to examine the validity of the measure, particularly after stakes are attached. Part of the evaluation process should be to identify and check for any unintended consequences that might occur once the indicator becomes part of the system. Examining whether and in what ways the indicator incentivizes schools toward actions that do not truly improve student outcomes is a large part of validating an indicator. An additional part of the evaluation process should be to seek ways to continue to refine the measure. For example, it is possible that changing how the metric is reported will improve the ability for schools and districts to take action based on the results.

CONCLUSION

As discussed in this paper, states have a variety of options for indicators, pathways, and measures available to them as they pursue innovative ways to comply with ESSA's 5th indicator provision. However, from conversations with state leaders, it seems that the many challenges of including new measures of school quality and student success appear insurmountable, leading several states to prioritize the use of chronic absenteeism as the sole 5th indicator. Chronic absenteeism can be collected and measured with relatively little additional burden on schools and districts, and it provides an important data point that fits ESSA's requirements.[47] However, a potential drawback of measuring student absenteeism alone is that it provides

46 Hough et al., 2016, p. 12.

47 http://www.hamiltonproject.org/assets/files/lessons_broadening_school_accountability_essa.pdf

limited information about what to do to improve student outcomes. To better address this, states may consider investigating approaches for multiple measures under the 5th indicator provision. This could take the form of examining measures that are already being collected by districts to better understand how to roll them up into a multi-metric indicator. It could also take the form of administering new measures through small-scale pilots, such as collecting student perceptions of school culture or social-emotional skills, in order to begin to examine the validity and reliability of such measures. Doing so has the added benefit of providing additional data for schools and districts for diagnostic and continuous improvement purposes. For ESSA to be truly used as an opportunity for experimentation and innovation, states need to be willing to study and assess indicators that move past NCLB and align with their goals and frameworks in order to better understand the evidence and think more deeply about feasible ways of incorporating new measures for future use in their accountability framework.

One Massachusetts Avenue, NW, Suite 700
Washington, DC 20001-1431
voice: 202.336.7000 | fax: 202.408.8072