In Search of Unicorns:

Conceptualizing and Validating the "Fifth Indicator" in ESSA Accountability Systems

Scott Marion and Susan Lyons

National Center for the Improvement of Educational Assessment

October 30, 2016

Abstract

This brief conceptual paper provides an overview of the accountability indicators required in the Every Student Succeeds Act (ESSA), focusing on the additional indicator of school quality or student success—the "fifth indicator." We provide a conceptualization of this indicator along two dimensions: 1) the information it adds to the accountability system; and 2) the level of inference. We recommend an indicator selection process that involves engaging stakeholders and developing a theory of action, and we present a framework for evaluating the intended indicators. We then highlight special considerations for establishing the weight for the additional indicator(s) within the entire accountability system and its role within the overall system design. The concluding section of this paper explains how the validity of the fifth indicator directly influences the validation of the ESSA accountability system as a whole.

Introduction

The Elementary and Secondary Education Act (ESEA) was reauthorized as the Every Student Succeeds Act (ESSA) in December 2015. The ESSA-required accountability system

must be operational in the 2017-2018 academic year. While 18 months may sound like a long time, a fairly quick design and development process is required to meet this timeframe. The law describes five types of indicators to be included in a school accountability system. First is the academic achievement indicator, also referred to as a status or point-in-time indicator. Under No Child Left Behind, achievement was reported as the percentage of students scoring at the proficient level or higher. Percent above cut (e.g., percent scoring at or above proficient) has been criticized for many measurement (e.g., reduction of information) and consequential (e.g., focusing on "bubble kids") reasons, but it does have the advantage of familiarity and relative ease of understanding and interpretability. While states are still required to report the percentage of students scoring proficient or higher, ESSA allows for approaches that rely on information gained from the full achievement distribution, such as an index system or average scale scores.

The second required indicator is another valid and reliable academic indicator. This indicator is required for elementary and middle schools and is optional for high schools. The law offers student growth and achievement gap closure as two potential options, but the indicator not limited to those examples. Measuring achievement gaps is one of the trickiest things to do well in educational measurement. Simple approaches such as computing the differences in percent proficient are almost always wrong, while more technically correct approaches such as computing the area between two performance distributions or effect sizes are a bit more challenging to explain to stakeholders. As difficult as it is to measure achievement gaps at any point in time, the challenges associated with measuring changes in achievement gaps are enormous. On the other hand, there are well-established methods for documenting student growth such as student growth percentiles (SGP), value-added models, and value tables.

Third, graduation rate must be part of the accountability system for high schools. Extended graduation rates such as five and six year rates can be included at the state's discretion. The fourth indicator is English language proficiency and progress, which became a new Title 1 accountability requirement when ESSA rolled Title III accountability into Title I. One of the key tenets of accountability design is that the application of the accountability rules should not systematically privilege or disadvantage schools based on demographics. Given that English language proficiency is typically relevant as an indicator in just a portion of schools within any state, accountability designers are going to have to work thoughtfully to ensure that schools responsible for developing English language proficiency in their students are held accountable without automatically disadvantaging the schools in accountability determinations.

For the last and "fifth indicator" listed in the law, ESSA requires the use of an additional indicator of school quality or success that meaningfully differentiates and is valid, reliable, and comparable. It is clear that the authors of ESSA wanted to broaden notions of school quality by including indicators in the system other than those based on test scores. The purpose of this paper is to provide a conceptualization of the fifth indicator, recommended a fifth indicator selection process, and highlight special considerations for establishing the weight for the fifth indicator(s) within the entire accountability system and determining its role within the overall system design. The concluding section of this paper is dedicated to discussing how the validity of the fifth indicator directly influences the validation of the ESSA accountability system as a whole.

Most of the indicators required under ESSA are at least familiar, even if the specific metrics proposed are new. However, the types of metrics and indicators suggested for the fifth indicator generally have not been used in accountability systems. Our colleagues sometimes

refer to this as the "unicorn indicator" because it is something we've all heard about, but rarely (or never) have seen. The specific passage from the law defining this indicator follows:

(B) INDICATORS.—
(v)(I) For all public schools in the State, not less than one indicator of school quality or success that—
(aa) allows for meaningful differentiation in school performance;
(bb) is valid, reliable, comparable, and statewide (with the same indicator or indicators used for each grade span, as such term is determined by the State); and
(cc) may include one or more of the measures described in subclause (II).
(II) For purposes of subclause (I), the State may include measures of—
(III) student engagement;
(IV) educator engagement;
(V) student access to and completion of advanced coursework;
(VI) postsecondary readiness;
(VII) school climate and safety; and
(VIII) any other indicator the State chooses that meets the requirements of this clause" (Every Student Succeeds Act, 2015).

As can be seen above, there are several psychometric characteristics required of this indicator. It must be valid, reliable, and must differentiate performance. But, in general, the options for what can be used as an indicator are fairly wide open. The following section of this paper provides a framework that may be useful for conceptualizing the characteristics that could define the types of metrics that would satisfy the requirements of the school quality or students success indicator.

Conceptualizing the Other Measure of School Quality

States need to be thoughtful about how this additional indicator fits with their conceptions of school quality and the roles and purposes of accountability within the state educational system. Before the process of selecting an indicator or set of indicators can begin, states must first articulate the rationale for including an additional indicator of school quality or student success within their accountability system. What role will the indicator play in realizing

the state's educational vision? For example, do education leaders think this additional indicator will serve to broaden the construct of school quality because previous test-based accountability systems have missed important aspects of school effectiveness? On the other hand, do leaders consider these indicators useful for accountability systems because they serve as additional measures of student success beyond academic achievement and growth? For example, some might want to include an indicator of student engagement because they think it is a hallmark of a high quality school and a necessary precursor to higher levels of student achievement. Or, could it be that states want to redefine success so that certain social-emotional learning indicators help broaden our understanding of what constitutes a successful student? We argue that this indicator can be characterized along the following two dimensions: information provided to system and level of inference.

Information Gained from the Fifth Indicator

Before operationalizing the fifth indicator, states must first articulate what information the inclusion of this indicator will add to the system. There are at least three ways to categorize indicators based on the information they provide: 1) identifying precursor indicators for achievement, 2) measuring additional aspects of the construct of school quality, or 3) measuring additional aspects of the construct of student success. Though there may be overlap across classifications, this framework helps conceptualize how this indicator will support the state's educational vision and theory of action.

Precursor indicators would be included in the accountability system because gains on these variables are thought to lead to greater gains in student achievement. This is currently the only type of indicator allowed under the draft regulations released by the United States Department of Education. According to the proposed regulations, there must be a research base indicating a link between success on the fifth indicator and increased academic achievement or

graduation rates. One might think about this category of indicators as leading indicators of future success. The second category is school quality indicators. School quality indicators are intended to broaden the ways in which we characterize school effectiveness beyond the typical indicators of achievement on standardized assessments in reading, mathematics, and science. They expand our measurement to include additional aspects of school quality that are not captured by student achievement. In some cases, school quality indicators can also be precursor variables in that they could identify particular mechanisms by which the accountability system creates and promotes improved student success. In that way, states could use a school quality variable to incentivize a particular best practice or set of practices for school improvement. Lastly, student success indicators broaden our definition of a successful student beyond achievement in reading, mathematics, and science. The use of these indicators underscores the idea that there should be multiple opportunities for students to demonstrate success, academic and otherwise. For example, students could be high achieving in competitive sports or the performing arts. This indicator would allow for the inclusion of such indicators in a school accountability system that values a definition of success beyond academic achievement. Table 1 summarizes the distinctions among the three categories. To help illustrate these differences, we offer several examples of potential indicators in each of the categories in Table 2.

Table 1
Conceptualizing fifth indicators

Pre-cursor Variables	Indicators of School Quality	Indicators of Student Success
• Causal precursors to the student success indicators. • Research base supporting link between improvement on these variables and improvement in academic achievement or graduation rates.	• Value propositions about what characterizes a high quality school. Broadens definition of school quality. • Process- or input-based • Mechanisms by which the accountability system will create and promote improved student success • Incentivize best practices in schools (specific change agents)	• Value propositions about what characterizes a successful student. Broadens definition of successful student. • Outcomes-based • Multiple viable opportunities for achieving student success

Table 2
Examples of fifth indicators

Pre-cursor Variables	Indicators of School Quality	Indicators of Student Success
• Algebra readiness by the end of 7th grade • Credits earned by end of ninth grade • Enrollment in advanced coursework • Chronic absenteeism	• Student/community engagement • School climate • Percentage of students in extra-curricular activities • Percentage of students enrolled in an art course • Educator quality (qualifications, experience, effectiveness) • Quality of local assessments or assessment practices • Engagement in professional capacity building	• Data drawn from post-secondary outcomes • Social-emotional skills • Physical fitness assessment results • Earning a career/technology certificate • Earning college credit • Persistence in post-secondary education • Entering STEM field

Level of inference

The level of inference associated with the indicator is an important dimension. A high inference indicator might be something like school climate, in which data are collected from students, educators, parents, and perhaps other stakeholders, usually through surveys and/or interviews. Once the data are collected, they are often transformed into scales that are thought to

relate to a construct of school climate. There are multiple steps along this inferential chain that must be validated to substantiate the claim that one is indeed measuring school climate. Likewise, student engagement also might be an indicator that falls along the higher inference end of the continuum; even if it is operationalized with something more straightforward like attendance. Nevertheless, strong inferences are required to support the notion that attendance is a reasonable proxy of engagement.

At the other end of the inferential continuum, we have indicators that basically rely on counting, such as counting the credits earned by the end of 9th grade or counting the number of students who have failed one or more courses in 9th grade. Of course, there is always some room for interpretation, such as determining what really counts as a credit for determining whether a student is on track for graduation. But as long as there is agreement on the business rules, this is still a low-inference indicator; this precursor indicator is not purporting to measure anything beyond credits earned.

Both high and low inference indicators have their benefits and drawbacks. In selecting an indicator or set of indicators, the top priority is defining what it is you would like to represent within your accountability system and your rationale for inclusion. Once the selected construct has been defined (e.g., educator quality, achievement gap), then the challenge becomes operationalizing that construct via high and/or low inference indicators. High inference indicators may lead to a more complete construct representation, but also typically require a heavier burden of evidence to support their validity (e.g., measures of social-emotional skills). On the other hand, low inference indicators are generally more easily calculated and provide for a more intuitive interpretation. However, they may be more distally related to the rationale for including the indicator (e.g., using AP enrollment to indicate college readiness).

Fifth Indicator Selection Process

Engage Stakeholders

As early in the process as possible, state educational leaders should convene a task force of stakeholders from across the state to be part of the planning and design process for the new ESSA accountability system. Key stakeholder groups that should be represented in every step of the indicator selection process include superintendents, principals, teachers, employers, high education representatives, advocates for civil rights (including advocacy groups for vulnerable children), parents and members of the public. Each of these groups brings a unique perspective and understanding about how the accountability system can function to bring about the intended change. Bringing together the voices of those who will be held accountable alongside those who the education and accountability systems are designed to serve can lead to new understandings and, ultimately, viable design options.

Articulate a Theory of Action

Articulating a clear theory of action that specifies the mechanisms and assumptions underlying how the accountability system is intended to work is a critical first step in designing the system. Effective accountability systems need to be designed according to a well-articulated theory of action that clearly lays out intended goals and outcomes, the proximal and intermediate indicators of progress towards those goals, and lastly the mechanisms and assumptions necessary for the system to function as intended. The theory of action becomes especially critical when selecting and ultimately validating an indicator or set of indicators to satisfy the fifth indicator requirement. Accountability systems should incentivize the types of behaviors state and district leaders want to see while discouraging counterproductive practices. The fifth indicator needs to be thought about in this light. Hargreaves and Braun (2013) offered concrete suggestions for

designing improvement-based (compared with punitive) accountability systems. While the design of the entire system should attend to these recommendations, the selection of the fifth indicator (or indicators) offers new opportunities to try to meet these improvement goals.

Part of the thinking about the theory of action for an accountability system is that it is both constrained and informed by the political, educational, and financial context in which it sits. Therefore, states should be considering using the ESSA accountability system to support behaviors aligned with larger system goals. States should view this fifth indicator as an opportunity to further important state policy goals rather than as a burden of "just one more thing" to include in the accountability system. For example, if a key goal is to promote access to higher education, a pertinent fifth indicator could be the percentage of students who have enrolled in and received credit for dual enrollment and/or Advanced Placement (AP) classes. Student and educator engagement may be distally related to the goal, but these indicators may be too indirect to incentivize the increased use of rigorous coursework in high schools around college preparation that states may hope to see on a large scale. This example is intended to illustrate how fifth indicators should be tied to supporting state-articulated goals, and through inclusion in the accountability system thereby increase the probability of achieving the stated goals.

Identify a Set of Indicators

Once the state goals have been articulated within a theory of action, the indicators that support the articulated intermediate and long-term goals can be identified. This step in the selection process involves identifying indicators suitable for serving their intended purposes, and it is likely that there may be many potential indicators that would fit the bill. As an example, it may be that the state has a goal to increase the percentage of adults with a high school diploma.

The theory of action for achieving this goal could involve statewide and targeted efforts to increase student engagement. Indicators that might be coherent with this theory of action could include surveys of student engagement or rates of chronic absenteeism.

Alternatively, the theory of action for achieving the stated goal of increased diploma attainment could instead involve increasing outreach and access to alternative diploma degree programs. This theory of action could call for indicators that broaden the definition of a high quality school to one that offers GED training courses that are open to the community. As illustrated by this example, the indicators or sets of indicators chosen will not only need to be reflective of the system goals, but also the theory of action for how the state intends for the goal to be reached. At this point in the process, it may be desirable to "over-generate" ideas of indicators that could be suitable within the system as all indicators are not likely to survive the evaluation process.

Evaluate Potential Fifth Indicators

Technical quality. After identifying a set of potential indicators and articulating the role they will play in realizing the state goals, it is important to evaluate the compliance of the potential indicators to ESSA requirements for technical quality, including validity, reliability (related to meaningful differentiation), and comparability. High-level questions for state consideration on these technical criteria are included in Table 3. Because the indicators states are considering for satisfying the fifth indicator requirement may be new to the accountability model, there is an increased responsibility or burden of evidence necessary to validate their use.

Table 3.
Technical Quality Considerations

Validity	• What theoretical and empirical evidence exists to support accuracy of the indicator as a measure of the underlying trait? • What is the relationship between the indicator(s) and school quality or student success? • How sensitive is the indicator to changes in school quality or student success (i.e., is the indicator malleable)? • What evidence supports the effectiveness of this indicator for achieving the intended outcomes in high-stakes accountability systems?
Reliability & Meaningful Differentiation	• How reliable is this indicator for making distinctions among schools within and across years? • What is the internal consistency of the indicator for measuring the intended construct?
Comparability & Fairness	• Is the indicator of school quality or student success equally representative of those constructs for all schools in the state? • Is the construct being measured in the same way across all schools? • Are comparisons among schools on the basis of this indicator appropriate and fair?

Data burden. In our zest to identify and collect data to improve our accountability determinations, we often forget that schools and districts are almost always short-handed and overwhelmed when it comes to new data collections. Indicator data that could be gleaned unobtrusively from records already submitted by districts or collected by the state would be the low end on the data burden continuum. A new data collection, such as school climate surveys (if not already collected), would probably fall somewhere in the middle of the continuum. Collecting data on real measures of student engagement that might require interviews and observations of students could be on the higher end of the data burden continuum. Unfortunately, given the structure of many school districts, organizing and submitting such data often falls to temporary workers or workers with limited training, raising some data quality concerns. We are not necessarily advocating for considering only measures with low data burden, but argue that such burden needs to be considered carefully.

In considering the data burden, it may be worthwhile to investigate sources of data that are already available to the state. The goal is to create an accountability system that is coherent with the broader state vision of what characterizes high quality schools. In the interests of both coherence and data accessibility, states may want to consider what data are already being collected that could be leveraged for the school quality indicator(s). For example, ESSA requires that states collect data to show that low-income and minority students are not served at disproportionate rates by ineffective, out-of-field, or inexperienced teachers. This kind of information could be aggregated at the school level to provide for an indicator of teaching quality in school accountability. Additionally, the state is collecting data pursuant to section 203(c)(1) of the Department of Education Organization Act, which may include rates of suspensions, expulsions, absenteeism, violence, and enrollment rates in accelerated coursework that could be used to inform indicators within the school accountability system.

Compliance. The Notice of Proposed Rule Making for ESSA accountability requirements would further necessitate that each indicator within the accountability system be disaggregated for each subgroup of students. Though this requirement directly contradicts best practice for survey-based measures such as the school climate and safety surveys offered as examples for this indicator in the language of ESSA, if the regulatory language is not changed, states must comply. This rule would exclude some measures of school quality that are collected or measured at the school-level (e.g., teacher engagement) and measures that that would require anonymous or confidential data collection to ensure the validity of the measure. For example, the identity of students responding to a survey about school climate should not be revealed to promote trustworthy feedback. Disaggregating by subgroup would require collecting respondent information and may compromise confidentiality and may have adverse impacts on

the accuracy of the measurement of certain constructs (see Stereotype Threat, e.g., Steele & Aronson, 1995).

Corruptibility. The indicators listed as examples in the statute and in Table 2 could provide rich information to schools and districts beyond test scores. However, many of the potential indicators such as school climate, student or teacher engagement, or other social-emotional indicators are often based on self-reported information gathered through surveys or other similar approaches. We must carefully consider "Campbell's Law" when using any indicator, but especially one that is easily corruptible if it used as part of a high stakes (or at least publicly reported) accountability system: "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (Campbell, 1979). While almost 40 years old, the truth of this "law" has been well documented since 1979, especially in the past 15 years. The double-edge sword described by Campbell is that not only may the indicator be corrupted, but the underlying trait or quality we are trying to measure will be distorted as well. In other words, we need to be very careful in our accountability system design and especially in the design of this fifth indicator.

One of the ways we may minimize corruption pressures is to consider multiple indicators for this category. For example, if this indicator was worth 15-20% of the overall rating, by using three to four indicators, each one would be worth only 5% of the overall score, which would lessen the risk of corruption because the potential reward is so small. Another way would be to consider indicators, such as some of the lower inference indicators discussed above, that require clear demonstrations of evidence to minimize corruption. Regardless of the composition and inference level of the indicator or indicators chosen, steps should be taken to closely monitor the

data quality over time. With this and other components of technical quality in mind, it is recommended that states first pilot any potential fifth indicators in a low-stakes, reporting-only capacity to study the nature of the indicator and establish baseline data before graduating the indicator(s) into the full accountability system.

Other unintended consequences. In addition to the corruptibility of the measure, states must consider the possibility of other kinds of unintended, negative consequences with the use of the chosen indicator(s) within a high-stakes school accountability system. Potential negative consequences can be identified and prioritized by revisiting the network of assumptions that underlies the theory of action for the accountability system. Through this process, we will find that some assumptions are more likely to be violated than others. In the event we identify an assumption that is either likely to be violated—or, if violated, consequences would be dire—this will signal a need to potentially flag the indicator for reconsideration or revise the theory of action (Hall, 2015). For example, imagine a state goal is to increase the rate at which students are graduating high school prepared for non-remedial college-level work. The state's theory of action for achieving this goal is to increase access to and success in rigorous, college-level coursework at the high school level. This state may consider including performance on Advanced Placement exams within the high school accountability system. However, the implicit assumption is that schools are offering Advanced Placement coursework to all students who are likely to benefit. A potential negative unintended consequence could occur if schools only allow students who are most likely to score 3 or higher on the AP to enroll in the courses instead of encouraging more students to take these rigorous courses. Therefore, states must consider the indicators in conjunction with the assumptions about their intended functions within the theory of action.

Considerations for the Weight of the Fifth Indicator

ESSA provides general guidance related to the weight of the fifth indicator within the accountability system. In particular, the first four indicators listed within the law must each be given "substantial weight," and "in the aggregate, much greater weight," than the fifth indicator. It seems that the lawmakers' intention is to prevent states from using the fifth indicator as a free pass to get out of the test-based accountability requirements. Though the requirements are clear, there is still a fair amount of variability in the weighting of the fifth indicator that would be allowable. Figure 1 shows three example of weighting schemes representing a range of possibilities for the fifth indicator in a high school accountability system that includes all five indicators.

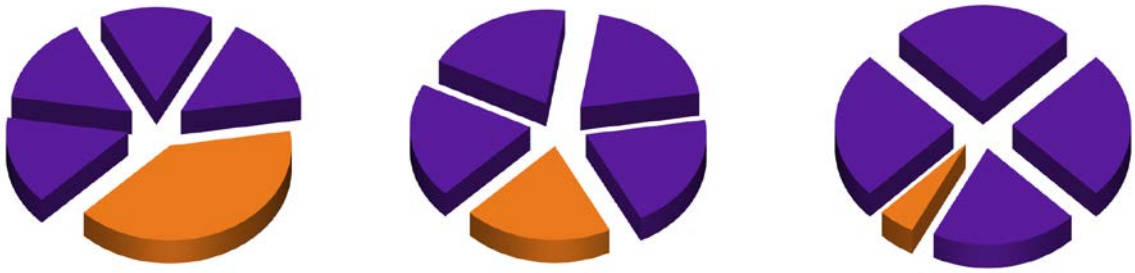

Figure 1. Example weighting schemata for fifth indicator

The weighting decision depends on both the confidence the state has in the technical quality of the indicator (e.g., the strength of the validity evidence to support its use) and also the policy-dictated "value" of the indicator. If there is little or no existing validity evidence to support the use of a desired fifth indicator, a state might want to first pilot the indicator within the low-stakes state reporting system before including it within the full accountability model. Another option would be to assign such a low weight within the system that it has no meaningful impact for the first few years while evidence of the technical quality is gathered. The policy-dictated value for the indicator will depend on the role of this indicator within the theory of

action. For example, does the theory of action require that this indicator to meaningfully contributes to the placement of schools into performance categories (higher weight), or does it only matter that the state signals to schools that indicator is valued (lower weight)?

The Notice of Proposed Rule Making would additionally require that the "School Quality or Student Success indicator(s) may not be used to change the identity of schools that would otherwise be identified for comprehensive support and improvement, unless such schools are making significant progress for the all students group under proposed §200.16(a)(1) on at least one of the indicators that is afforded substantial weight and can be measured for all students" (Every Student Succeeds Act NPRM, 2016). We argue that this requirement excessively diminishes the value of the school quality or student success indicator and may erode motivation to focus on an important area of improvement. Additionally, this requirement would remove a legitimate basis for meaningfully differentiating between schools that might otherwise be regarded as similarly performing. We believe the requirements outlined in ESSA provide enough guidance to ensure the school quality or student success indicator does not account for undue weight; however, if these rules are upheld, they will need to be factored into the weighting decisions and likely require extensive data modelling to ensure compliance.

Decisions regarding how much any indicator is weighted within an accountability system can only be made once all the indicators for inclusion have been identified, and ideally, once pilot or extant data has been gathered. Because the effective weight of an indicator within a composite score is not only dictated by its policy weight but also by the amount of unique variability it contributes to the composite, empirical analyses are an essential part of evaluating the success of the policy-dictated weighting scheme at generating the intended summative accountability score.

Conclusion

The validity of the selected fifth indicator(s) rests on both the degree to which they measure what they purport to measure and the appropriateness of their use within the larger accountability system (AERA, APA & NCME, 2014). Not only must the validity evidence support the technical quality of the indicator itself, but it must also support the use of the indicator for its intended purposes. In accountability systems, the purpose for including the fifth indicator will often be as a mechanism for change as outlined in the theory of action. Therefore, the framework presented in this paper for the fifth indicator can be used to not only guide the approach for indicator selection, but also the validation. For example, if the indicator represents something distinct from traditional test-based academic achievement, then we would not necessarily expect a strong relationship between assessment results and favorable performance on this indicator. One might think of a community engagement initiative that encourages students to participate in service activities or other applied projects. Such engagement may be thought to help students hone leadership skills and other characteristics associated with being responsible global citizens, both of which are not measured well on tests. It stands to reason, then, that validating the indicator with evidence of correlation with traditional assessment data would be misplaced. Rather, we would seek other data thought to affirm our understanding of the construct. On the other hand, a state's theory of action may hypothesize that encouraging students to be engaged in community service or other applied projects increases motivation and hones critical thinking skills essential to academic success. With this view, one expects students who are more engaged to perform better on academic assessments. If not, our understanding of the construct and the role the indicator is intended to play within the accountability is less certain. Importantly, these contrasting conceptions of student engagement are both potential

hypotheses that should be clearly articulated in a theory of action and then evaluated empirically as data are collected.

The inclusion of an additional indicator of school quality or student success will likely be new territory for many states engaged in accountability design in the coming months and years. As states select and validate new kinds of indicators within their accountability systems, we encourage them to publish and share their theories of action alongside the empirical and theoretical support for the fifth indicator and accountability system as a whole so all states can benefit from the innovation that is likely to occur as a result of this new legislation.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*(1), 67-90.

Every Student Succeeds Act §1111, S.1177. (2015, December 10). 114th Congress. Retrieved from: https://www.congress.gov/bill/114th-congress/senate-bill/1177/text

Every Student Succeeds Act §1111, S.1177. Notice of Proposed Rule Making. (2016, May 31). 81 Federal Register 34539, 34539-34621. Retrieved from: https://www.federalregister.gov/documents/2016/05/31/2016-12451/elementary-and-secondary-education-act-of-1965-as-amended-by-the-every-student-succeeds

Hall, E. (2015, October). *Using a theory of action to support the identification of negative consequences related to the use of test scores in educator evaluation*. Paper presented at the 2015 annual meeting of the New England Research Association, Trumbull, CT.

Hargreaves, A. & Braun, H. (2013). *Data-Driven Improvement and Accountability.* Boulder, CO: National Education Policy Center. Retrieved September 14, 2016 from http://education.nh.gov/essa/documents/data-driven-improvement.pdf

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology, 69*(5), 797.