New Hampshire
Department of Education

Every Student Succeeds Act (ESSA) Accountability Task Force Meeting

July 7, 2016
NH DOE Room 15

AGENDA

OBJECTIVES FOR MEETING:

1. Sharing and discussing elaborated potential ESSA accountability framework.
2. Determining how to translate the NH Coalition for Business and Education goal of having 65% of 25-64 year olds with a high-quality postsecondary credential by 2025 into ESSA accountability goals and intermediate targets.
3. Continue conceptualizing and nominating potential "additional indicators of school quality."

1:00 Welcome and policy updates

- ✓ Virginia Barry and Paul Leather, NH DOE

1:15 Framing a potential NH Accountability System Design

- ✓ The slide deck presented at the last meeting has been converted to a brief document to summarize the discussion of the slides and to add more detail than could be presented in the slides. Our discussion this afternoon is designed to solicit input on the direction of the model and to shape subsequent model decisions.
  - **Document**: NH ESSA Accountability Framework

2:15 65% x 2025

- ✓ At the last meeting the Task Force recommended capitalizing on the political and stakeholder buy-in to a long term goal established by the NH Coalition for Business and Education as part of a strategy to ensure the sustainability of NH's workforce. This goal is to have 65% of 25-64 year olds possess a high-quality postsecondary credential by 2025. Our challenge will be to "walk this back" so that we can translate this first to a K-12 goal in 2025 and then an accountability target in 2018 through 2025.
  - **Slide deck**: 65 x 25 Final Recommendations (HCM Strategists)
  - Goal setting activity

3:00 Other indicators of school quality

- ✓ We started discussing the other measures of school quality (aka 5th indicator) at our last meeting. We offered several ways to conceptualize these measures that are explained in more detail in the "5th Indicator Brief" provided with this agenda.
  - How are we conceptualizing our additional measure of school quality?
    - School quality or student success
    - Low or high inference

- Low or high data burden
- Low or high potential corruptibility

3:55 Next steps and next meetings

- ✓ August 16 9:00-Noon
- ✓ September 9 9:00-Noon
- ✓ October 4 9:00-Noon
- ✓ November 2 1:00-4:00
- ✓ December 8 9:00-Noon

4:00 Adjourn

Potential Career-readiness Student Metrics:

1. Number and percentage of all students with access to career pathways
    a. in high skill, high demand sectors.

2. Number and percentage of all students who complete career pathways in
    a. high skill, high demand sectors.

3. Number and percentage of all students who earn college credit that transfers to a higher education institution

4. Number and percentage of all students who earn industry recognized credentials (ICRs)
    a. in high skill, high demand sectors.

5. Number and percentage of all students who enroll in college or secure employment or who are enrolled in college within 12 and 18 months of high school graduation
    a. in high skill high demand sectors

6. Number and percentage of students who participate in a work-based learning experiences
    a. in a high skill, high demand sector.

7. Number and percentage of students who possess work-ready employability skills

Other Potential Additional Academic Indicators

<u>Existing Indicators</u>
Attendance
School Approval
Teacher Evaluation
Class Size
YRBS
School climate
School Safety
Graduation rate
Dropout rate

<u>Other potential indicators</u>
Career Readiness
Student engagement
Enrollment in advanced coursework
School climate surveys
Participation in extra-curricular activities

Percentage of students enrolled in an art course
Educator quality (qualifications, experience, effectiveness)
Suspensions/expulsions
Quality of local assessments or assessment practices
Engagement in professional capacity building
Achievement gap indicator
Persistence
Data drawn from post-secondary outcomes
Social-emotional skills
Physical fitness assessment results
Credits earned by end of ninth grade
Algebra readiness by end of 7th grade
Access/completion of advanced coursework
Percentage of students entering STEM field
Postsecondary readiness
Persistence in post-secondary education

Goal-Setting for the New Hampshire Accountability Task Force

As a result of the New Hampshire Business and Education Coalition's stated goal of 65% of students ready by 2025, the New Hampshire Accountability Task Force will need to consider how to expand upon that goal for New Hampshire students. This will require identifying the indicators and determining how progress on those indicators contributes to the accountability system. This must be done within the context of what college- and career-ready means.

The concept of college- and career-ready is often split between distinct measures of college **or** career ready. This notion may be further supported by the idea that many of the career-ready measures that states are using reflect performance for those students who may not have plans to attend a 2- or 4-year college. In support of this, it may be most appropriate for accountability systems to incorporate disjunctive systems that allow students with different strengths and skillsets to demonstrate readiness in the way in which is most appropriate for them (their path? their educational experience).

Potential Indicators

As the task force considers the most appropriate indicators, it is important to distinguish between those students who are likely to attend 2- or 4-year post-secondary institutions and those who will likely go straight into employment. The following indicators are examples of what may be used to determine readiness.

1. College-ready measures
    a. Performance on the state summative assessment
    b. Predictive score on the SAT (or ACT)
    c. 3+ on Advanced Placement exam
    d. 4+ on International Baccalaureate exam
    e. Grade to be determined (e.g., B or higher) on Department approved dual enrollment courses
2. Career-ready measures (note, these are potential examples)
    a. In conjunction with a defined score on the state standardized assessment
        i. Industry recognized certificate
        ii. Credential earned through an approved CTE program of study that holds value professionally, for post-secondary enrollment, or in an associate or baccalaureate degree program
    b. As a standalone indicator of readiness
        i. Anything in a above with completion of a co-op job training opportunity
        ii. X (e.g., 2) number of workplace experiences "courses" in any area (note: if NH has these at all)
        iii. X number (e.g., 2) of successfully completed CTE courses in pre-defined categories by the department (note: these is less stringent, obviously)

Evaluating Performance Standards for Indicators

Once indicators are selected, the Accountability Task Force must also determine how performance standards for each indicator are defined. If the task force determines that a disjunctive approach is appropriate, the calculation is relatively simple. It can be displayed as:

$$\frac{\text{\# of students who meet any one of the criteria above by the end of 12}^{\text{th}}\text{ grade}}{\text{Total \# of students enrolled}}$$

With the end goal of 65% by 2025 known, the Task Force should examine at least 2 pieces of data to determine how difficult it may be to close the distance to the target. The first sets of data will likely be the easiest to obtain, but both can be useful. These include the following:

1. Historical trend data on any of the previous indicators to inform the degree to which adjustments will be required to maintain or accelerate improvement over time; and
2. Any available data that links high-school performance on each of the indicators to post-secondary data. This connection can help validate whether proxies in high school reflect success in college or careers.

If the second source of data is not available, the first can at least inform the reasonableness of the requirement improvement between current performance and the 65% target in 2025.

65x25 – Update

- Goal is for 65% to have some credentials by 2025 for **25 to 64 year olds.** What does this look like in k-12?
    - What does that look like for exiting 18 year-olds?
    - What does that look like for 9th-11th graders?
    - What does that look like for middle schools for 5th-8th graders?
- The answer to these questions essentially defines the indicators that are feasible to use. The high school indicators are likely the high-stakes versions, but the middle school (and then elementary school) indicators could be used as low-stakes or precursor indicators of success.

What does the process look like? If we can get buy-in on the process from the task force, along with an identification of groups of college and career indicators, a lot of headway could be made between now and the August task force meeting. What does said process look like?

- We have 3 equations but multiple knowns
    - 25-64 year old credentialing
    - 18-24 credentialing (degree obtainment & career placement)
    - 18 and under school performance
- What are the knowns?
    - 25-64 credentialing: 53% now, but we need to get to 65%
    - 18-24 credentialing:
        - "college-goers"
            - What percentages of NH students enter 2 or 4-year colleges without needing remediation?
            - What % graduate within 4 or 6 years from 2 and 4 year institutions?
        - "career-goers"
            - What % are successfully placed in occupations with industry-standard wages?
            - What % receive some occupational certification
            - What % receive some occupational apprenticeship?
    - 18 and under school performance (from previous documents)
        - College-ready measures
            - Performance on the state summative assessment
            - Predictive score on the SAT (or ACT)
            - 3+ on Advanced Placement exam
            - 4+ on International Baccalaureate exam
            - Grade to be determined (e.g., B or higher) on Department approved dual enrollment courses
        - Career-ready measures (note, these are potential examples)
            - In conjunction with a defined score on the state standardized assessment

        - Industry recognized certificate
        - Credential earned through an approved CTE program of study that holds value professionally, for post-secondary enrollment, or in an associate or baccalaureate degree program
    - As a standalone indicator of readiness
        - Anything in a above with completion of a co-op job training opportunity
        - X (e.g., 2) number of workplace experiences "courses" in any area (note: if NH has these at all)
        - X number (e.g., 2) of successfully completed CTE courses in pre-defined categories by the department (note: these is less stringent, obviously)
- What gets us to 53%? Need to determine what the indicators in HSs look like
    - If the indicators are defined in high school, then the two sets of indicators can be compared
    - We can identify what the current data look like, but need to understand what the historical trends are as well
    - We can then also identify the differences in rates of change

Every Student Succeeds Act
Considering the Impact of CCR Indicators and Design Criteria for Accountability Systems

Introduction

In December 2015, the Elementary and Secondary Education Act (ESEA) was reauthorized and signed into law as the Every Student Succeeds Act (ESSA), which requires all states to redesign their accountability systems by the 2017-18 school year. A strong, coherent accountability strategy that ties together graduation requirements, assessments, and other indicators of readiness is essential to provide information that can guide school, district, and state strategies for increasing college and career readiness. As states make changes to their accountability and public reporting systems under ESSA, they have an opportunity to select indicators that inform continuous improvement, guide decisions about resource allocation, inform policies, capture progress, and factor into accountability determinations. These indicators should also reflect a continuum of performance including toward, meeting, and exceeding readiness. Additionally, the flexibility provided by ESSA on the types of indicators that need to be included in accountability systems has led to significant interest in measuring, for example, indicators of school quality and success, and English language proficiency.

As states begin the work of transitioning to new ESSA accountability systems, Achieve and the National Center for Improvement of Educational Assessment have partnered to release this guidance brief for states. This brief seeks to address the following central questions:

> *What opportunities are afforded states under the ESSA and how might those opportunities advance states' policy goals?*
>
> *What is the impact of ESSA legislation and the additional indicators on accountability systems with regard to policy and technical criteria?*

Comment [JD1]: I still think there's a way to blend these two into a single question. It's beyond CCR but either way should factor in the criteria.

The opportunities under ESSA should use a set of policy and technical criteria that can help states design and evaluate their accountability systems. **Policy criteria** refer to aspects of the system that support its intended goals, purposes, and uses. **Technical criteria** refer to the extent to which the system produces valid and reliable classifications or ratings. Further, by framing the design in terms of a theory of action while addressing policy and technical criteria, practitioners can use an accountability system's goals and intended outcomes to inform how indicators and their design principles are leveraged. However, these goals must be considered within the bounds of ESSA. This brief presents a summary of accountability provisions under ESSA, policy and technical criteria to evaluate the effectiveness of accountability systems, and a set of realistic, hypothetical scenarios using different approaches to indicator selection and use. Through these, we can concretize the issues associated with the criteria raised and their impact on the design and implementation of accountability systems.

Every Student Succeeds Act

The ESSA-required accountability system must be **operational in the 2017-2018 school year**. While a theory of action may articulate explicit goals of the accountability system, it can likely be assumed that a state's accountability system will forward notions of program monitoring, school improvement, and providing school ratings. This timeline and these needs necessitate a fairly quick design and development process. There are two main components of the ESSA accountability system:

1. Reporting requirements: States must continue to report by all required subgroups specified under NCLB, as well as the additional subgroups of foster children, homeless students, and children of military families for assessment reporting and accountability indicator reporting.
2. School accountability determinations: States must categorize schools based on state-determined goals and methodology.

While ESSA provides an opportunity to consider new accountability systems, the recommendations presented in this paper are relevant when considering how one should design an accountability system in general. States are in varying positions with regard to the types of data and metrics that can be leveraged. For example, some states that have implemented accountability systems either based on NCLB waivers or not, may desire to make some minor adjustments or validate existing accountability systems that align with the spirit of ESSA. Other states may see this as an opportunity to restructure their accountability systems entirely. Regardless of the state's orientation, this paper offers tangible approaches to examine how the accountability system supports the theory of action set forward.

Accountability Indicators[1]

It is important to keep in mind that while ESSA outlines the basic structure of state accountability systems, the specifics of the accountability design may[2] be worked out in the rule making process. However, this paper does not go into great detail regarding the identification of targeted support or comprehensive support schools. **Targeted support schools** are those schools that have either (1) a low-performing subgroup that falls below the lowest-performing 5% of Title I schools in comprehensive support or (2) has at least one consistently underperforming subgroup based on state-defined criteria using no more than two years' data. **Comprehensive support schools** are those that include (1) the lowest-performing 5% of all Title I schools in the state, (2) any public high school failing to graduate one-third or more of its students, or (3) a chronically low-performing subgroup where a subgroup is performing as poorly as the lowest-performing 5% of Title I schools and has failed to improve after implementation of targeted support over no more than three years. While the current draft regulations offer some information on how these targeted and comprehensive support schools should be identified and supported, states may want to focus instead on the law itself until the regulations are finalized.

[1] From Marion, S. F. (2016). 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. Dover, NH. http://www.nciea.org/publication_PDFs/ESSA%20Accountability%20Design%20Considerations_021916.pdf

[2] We say "may" because ESSA's main sponsor, Senator Alexander, has been vocal in his opposition to the potential of overregulation of ESSA.

The law describes five types of indicators to be included in a school accountability system:

1. **Academic achievement** is also referred to as status or point-in-time indicators. Under the No Child Left behind Act (NCLB), achievement was reported as the percentage of students scoring at the proficient level or higher. Percent above cut (e.g., 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. While states are still required to report percent proficient, ESSA may allow for approaches that rely on information throughout the achievement distribution such as an index system (something familiar in many waiver states) or average (i.e., mean) scale scores.
2. **Another valid and reliable academic indicator** must be included in the accountability system. The law offers student growth and achievement gap closure as two potential examples, but it is not limited to those examples. While challenges exist with these indicators ranging from accuracy and precision to understandability and access, the familiarity and usefulness of indicators currently used in states may be an important factor in selecting appropriate measures for this category of indicator.
3. **Graduation rate** must be part of the accountability system for high schools. Further, extended graduation rates such as five- and six-year rates can be included at the state's discretion. However, the 4-year adjusted cohort graduation rate is required for the identification of comprehensive support schools.
4. **English language proficiency** rates and progress is a new accountability requirement under ESSA, at least under Title I accountability. This is largely because this part of Title III accountability has now been rolled into Title I. This is one of the aspects of ESSA that will need rules to help us better understand the requirements. For example, one of the key tenets of accountability design is that accountability rules should not privilege or penalize schools based on the demographic characteristics of the school. Given that the populations of English language learners are not randomly distributed across school districts in states, thoughtful design work will have to be made to minimize unintended consequences. That is, that schools that are responsible for developing English language proficiency in their students are held accountable, but that the presence of this indicator does not automatically disadvantage the school in accountability determinations.
5. ESSA also requires the use of an **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. Again, this indicator is dependent on the goals and intended outcomes of an accountability system, as well as those measures that are readily available within a state.

Design and Subgroup Considerations

One area of uncertainty in ESSA involves how subgroup performance must be considered in accountability designations and for targeted support and improvement. From initial listening sessions

with ED and the proposed ESSA regulations[3], how subgroups must be included in overall accountability determinations is still unclear. From early reads, it appears that individual subgroups, while playing an important role in reporting and certain types of identification (e.g., targeted and comprehensive support schools) may not be required to calculate accountability determinations. In addition to the selection of indicators and their combined use, subgroup performance will need to be considered.

The sections in ESSA regarding the inclusion and consideration of subgroups force a dual approach of both how the system components fit together and how the individual indicators are calculated. According to ESSA, subgroups must be considered in the following manner:

> (A)(i) Establish ambitious State-designed long-term goals, which shall include measurements of interim progress toward meeting such goals;
>
> (B) INDICATORS—Except for the indicator described in clause (iv), annually measure, for all students and separately for each subgroup of students, the following indicators…

This may imply that while the indicators are used for school differentiation, both the indicators and long-term goals disaggregated by subgroup may be primarily for reporting and transparency rather than high-stakes accountability (exclusive of targeted and comprehensive support). That is, there is a distinction between two potential sets of purposes of the accountability system. First, an accountability system and its disaggregated indicators should support setting goals and reporting performance. Second, the system should use outcomes to inform consequences, supports, ratings or classifications. The first set is of vital importance in reflecting the mechanisms and processes forwarded in a theory of action. Additionally, it can help provide guidance to schools, districts, and states using connected and relevant data to support school improvement. The second, or the high-stakes ratings or classifications, should be developed thoughtfully to send the appropriate signals aligned with rewards and supports.

Given that the proposed regulatory guidance has stipulated the use of subgroups for the classification of Targeted and Comprehensive support schools, agencies will need to determine how their school classification systems either align or thoughtfully separate themselves from those requirements. Importantly, agencies will need to be cognizant of how the regulatory guidance changes as a result of public comment and revision.

There are several ways that state leaders may consider how multiple measures are combined as part of the accountability system design. Five possible approaches are described below.

1. Decision matrix approach: This approach would use profiles of school performance that demonstrated certain levels of school quality. A committee would determine how profiles are established, any decisions rules related to the profiles, and guardrails and requirements for meeting expectations.
2. Index approach: This approach would use numerical aggregations of performance across indicators. This would require agencies to determine performance standards for each of the

[3]Notice of Proposed Rulemaking for the Every Student Succeeds Act (May 31, 2016). http://www2.ed.gov/policy/elsec/leg/essa/nprmaccountabilitystateplans52016.pdf

individual indicators as well as the aggregation rules across indicators. Index approaches may also include conjunctive rules, compensatory rules, or a combination of both.

3. Conjunctive approach: A conjunctive system is one that requires all indicator targets to be met (i.e., a series of AND statements) in order to meet the overall target. This case would be the most restrictive. Likely, each indicator would have an annual target, which would be applied to each subgroup that met the minimum n in a school.
4. Compensatory approach: A compensatory system is one that allows for higher performance in one indicator to compensate for lower performance in another, with some overall target being defined. Such a design should factor the ESSA requirement that the first four indicators have a much greater weight than the 5th indicator. Subgroups would be used for reporting and any other required interventions, but not for assigning school ratings. Alternatively, a compensatory system could be applied to the indicators and subgroups in order to increase alignment across the system.
5. Hybrid approach: This would be a combination of the conjunctive and compensatory approaches. Instead of a single target per indicator per subgroup (i.e., conjunctive), there may be a compensatory target for **each subgroup** using all 5 indicators. The conjunctivity would then be applied for overall subgroup information.

Each of these approaches affords a different degree of differentiation and carries different technical characteristics. Given the restrictive and potentially damaging outcomes associated with conjunctive approaches, states should consider opportunities to utilize measures and system designs that promote differentiation and diagnosis. However, safety triggers (e.g., conjunctive flags, minimum thresholds, etc.) should be thoughtfully considered to ensure instances of false positives and negatives (e.g., a "high performing" school is identified for improvement or a "poor performing" school is not identified for improvement because of the way the system has been designed) are avoided. In an effort to suggest a system that validly differentiates among various levels of "school quality," the remainder of this paper focuses on the criteria to evaluate the efficacy of accountability systems and the dimensions one should consider when evaluating these systems that align to a clearly articulated theory of action.

Facing Page Graphic?

Conjunctive Approach: Conjunctive for both indicators and subgroups							
Subgroup	**Achieve**	**Growth/ Gap**	**Grad Rate**	**ELP**	**5th**	**Overall**	**Notes (example rules)**
All	✓	✓	✓	X	✓	Yes	Target has been met on a majority of indicators
Low SES	X	✓	X	X	X	No	Target has not been met on a majority of indicators
Compensatory Approach: No conjunctivity							
All	50	60	75	60	80	**325**	Target has been met based on an initial baseline and eventual end goal using all available data. Fully compensatory, or compensatory with 5th indicator providing "bonus" points.

Each Subgroup							Here, each indicator could be disaggregated for a given subgroup to provide transparent reporting, but would not necessary be used for high stakes accountability decisions.
Hybrid Approach: Conjunctive across subgroups							
All	50	60	75	60	80	**325**	Compensatory target has been met
Low SES	40	55	70	NA	75	**240**	Compensatory target has not been met

Honoring Priorities using a Theory of Action

The selection of indicators for any accountability system is dependent in large part on a defensible theory of action. This includes clear articulation of what signals the accountability system is trying to communicate to the public, educators, administrators, and policy-makers.

With regard to an ESSA accountability system, the theory of action should describe the mechanisms that will bring about the state's clearly defined goals. That is, what actions, resources, processes, and supports must be put in place so that the state can achieve its goals? Furthermore, what evidence can and will the state collect to demonstrate that these actions lead to the expected goals? By proposing a theory of action that is based on empirical evidence and logical connections between steps, it can serve as a filter throughout the development process and ensure that practitioners are making sound assumptions.

Developing a Theory of Action

While there is no single approach used for creating a theory of action, several resources exist to help practitioners develop a theory of action (e.g., the Kellogg Foundation's Logic Model Development Guide[4] or Shakman & Rodriguez' Logic Model Toolkit[5]). Readers are also recommended to reference the Marion, Lyons and D'Brot (2016[6]) framework as an overview of a potentially useful process to develop a theory of action for ESSA-based accountability systems.

1. Clearly describe the **goals** of the accountability system.
2. Articulate the **purposes** and intended **uses** of the accountability system results.
3. Define the specific intended **outcomes** of the system.

[4] 4 W.K. Kellogg Foundation. (2006). *Logic model development guide*. Battle Creek, MI: Author. Retrieved April 13, 2013, from http://www.wkkf.org/knowledge-center/resources/ 2006/02/wk-kellogg-foundation-logic-model-development-guide.aspx
5 Shakman, K., & Rodriguez, S. M. (2015). *Logic models for program design, implementation, and evaluation: Workshop toolkit* (REL 2015–057). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs
[6] Marion, S. M., Lyons, S., D'Brot, J. (2016). Developing a theory of action to support high quality accountability system design. *National Center for the Improvement of Educational Assessment:* Dover, NH. http://www.nciea.org/publication_PDFs/ESSA%20Accountability%20Design%20Considerations_021916.pdf

4. Lay out the **mediating outcomes** necessary to achieve the ultimate outcome(s).

5. Create an initial "high-level" (large grain size) **theory of action** as a first step to mapping out the components.

6. Build off the "high-level" theory of action and **add** enough **details** to articulate how these major components relate to the minor components.

7. "Zoom-in" on several key components of the theory of action to add the detail necessary to support the **accountability design and the validity evaluation** as a final step.

Through the thoughtful design, construction, and articulation of a theory of action, practitioners will have a rich resource that can highlight which mechanisms or programs may be untenable, which will support the system's goals, and which may be missing. Further, the theory of action will serve as both a record and framework for guiding the evaluation of the proposed system. The theory of action allows evaluators to isolate aspects of the system that may not be working as intended instead of having to "throw the baby out with the bath water" because the ultimate goals may not have been met.

Criteria to Evaluate the Efficacy of Accountability Systems

Educational accountability systems should be designed to contribute to the improvement of educational systems[7]. However, there is also a need to define how the accountability system defines the quality of education and how it is communicated to stakeholders (e.g., educators, administrators, policy-makers, community members, and students). Both policy and technical factors should be considered when evaluating the efficacy of accountability systems. These include (1) policy considerations and (2) technical considerations.

Stated at the beginning of this paper, the reader is reminded of the central questions posed:

> *What opportunities are afforded states under the ESSA and how might those opportunities advance states' policy goals?*
>
> *What is the impact of ESSA legislation and the additional indicators on accountability systems with regard to policy and technical criteria?*

The policy and technical criteria serve as a set of filters one can use to consider specific metrics and measures. However, it can be easy to become consumed by these criteria and neglect the theory of action (ToA) upon which the accountability system should be designed. The ToA should be used as the basis of answering these questions. That is, what are the hypothesized mechanisms, processes, components, and programs that will bring about the clearly stated goals in support of an ESSA-based accountability system? If those mechanisms and components empirically and logically support a state's goals, one can begin to use the policy and technical criteria as a filter for appropriate consideration and

[7] Hargreaves, A. & Braun, H. (2013). *Data-Driven Improvement and Accountability.* Boulder, CO; National Education Policy Center. Retrieved April 13, 2016 from http://nepc.colorado.edu/publication/data-driven-improvement-accountability/.

application of available metrics and measures, thus establishing a validity argument for the accountability system.

While both are important guiding factors, it is possible that there may be a tension between the two. That is, an indicator may have a high policy value but may not perform well against technical criteria for inclusion in an accountability system. In this case, the theory of action should be used as a guide to determine whether the measure should be a part of the accountability system. The remainder of this section forwards a sample set of considerations and should not be taken as an exhaustive list. Rather, other criteria should also be included depending on the system and its components that are deemed important and relevant by the agency's stakeholders and practitioners.

Policy Criteria

Policy criteria are those factors that can be used to determine how well the *why* and *how* (i.e., the ToA) serve the intent of the accountability system. They can also be thought of as levers that seek to instantiate certain values and incentivize desired actions. These criteria include, but are not limited to, concepts such as overall goals, equity, promotion of college- and career-readiness, broadening the construct of school quality, increasing the depth of current measures of school quality, corruptibility, and utility. While only a partial list, these ideas are intended to highlight certain types of policy goals. We offer several policy considerations to use as examples below. They are explained in more detail in *Appendix X.*

- **Policy goals.** The overall policy goals set forth in the accountability system should be a primary driver in designing and evaluating an accountability system.
- **Equity.** Equity in accountability systems seek to incentivize actions that lead to academic improvement for the lowest performing students and in detecting desired outcomes (see Domaleski & Perie, 2012[8] for more detail on approaches to emphasize equity in accountability systems).
- **Promoting college- and career-readiness.** An example goal like this will require practitioners to consider what components and processes in the theory of action contribute to summative (i.e., end-of-process) indicators of CCR. This can be done potentially by using short- and mid-term indicators as part of low-stakes reporting that informs overall progress.
- **Focusing on other measures of school quality.** A potentially powerful aspect of ESSA is the inclusion of additional measures of school quality. One can conceptualize this as a means to broaden the construct of school quality, as a way to increase the depth of current measures of school quality, or as precursors to growth and achievement.
    - *Broadening the construct of school quality***.** This widens the view of school performance to include those indicators that could paint a more holistic picture of school quality.

[8] Domaleski, C. & Perie, M. (2012). *Promoting equity in state education accountability systems.* National Center for the Improvement of Educational Assessment. Dover, NH. http://www.nciea.org/publication_PDFs/ Promoting%20Equity%20CSDMP110712.pdf

    - *Increasing the depth of current academic measures of school quality***.** This widens the view of school quality by highlighting certain mechanisms that support academic success.
    - *Precursor indicators of growth and achievement.* This would consider those potential process indicators that inform our understanding of why schools and students might be exhibiting stagnant, declining, or increasing growth and achievement.
- **Resistance to corruptibility.** Corruptibility refers to the alteration or deterioration of inferences made from information due to intentional or unintentional changes in practice. While process and subjective measures may be more easily corrupted, states should be mindful that objective and standardized measures are also at risk.
- **Utility.** As accountability systems are put into place, states must examine how well the system supports the intended goals without deteriorating other aspects of the educational system. Does the accountability system serve the practical information and behavioral needs of users?

Policy statements often serve as proxies for the larger goals of the educational and/or accountability system. However, the policy considerations quickly get into the detail associated with the ToA. The policy-oriented goals of the accountability system should serve as the primary filter for the identification and selection of accountability measures. More specifically, practitioners should use the ToA to determine the degree to which measures align with the outcomes, broadness or tightness of focus, and long-term claims being made. By using the policy criteria as a filter in development, practitioners can more quickly identify measures that support the theory of action.

Technical Criteria

The technical criteria can then serve as a second filter when selecting measures for an accountability system. That is not to say that one should only consider the policy considerations. Rather, the technical considerations supplement the policy considerations. The technical set serves to help practitioners amass evidence that the selected measures meet the goals of the system and perform satisfactorily. By leveraging the technical set, one can examine the way in which measures function on past and current datasets to ferret out any problems that may arise in operational use. Further, the technical considerations can be revisited once the accountability system is in operation to gather evidence that the system is working as intended. In other words, the technical criteria may be necessary, but not sufficient, to achieve the intended goals of the accountability system.

The technical criteria include concepts such as technical goals, availability, reliability, comparability, and equity. While only a partial list, these ideas are intended to highlight certain types of technical goals. We offer several technical considerations to use as examples below. They are explained in more detail in *Appendix X.*

- **Technical goals.** In the same way that the theory of action and the accountability system forward a set of policy goals, they should also include technical goals selected to fit the policy objectives.

- **Availability.** To what extent are data elements available across grade spans and schools within a state?
- **Comparability.** This criterion raises the question of how well comparisons can be made from site to site.
- **Reliability.** Reliability is a framework for quantifying the uncertainty associated with sampling whether from a population (e.g., students, schools) or universe of possible measures (e.g., items, tests, survey questions) and can be viewed as the consistency across replications of a certain event as quantified using a variety of metrics[9], (e.g., reliability indices, measurement error, sampling error, etc.). However, readers are encouraged to review Hill and DePascale's (2002[10]) discussion on methods to determine the reliability of an accountability system. In this application, reliability can address the consistency of the metrics being used (and how they are combined) and the consistency of the population.
- **Equity.** Initially introduced as policy criteria, equity serves to empirically determine whether the ideas of fairness and inclusion are upheld. This may be done by determining whether a relationship exists where there should be none (e.g., between the measures in an accountability system and characteristics of the population).
- **Focusing on other Measures of School Quality.** While this was also included as part of the policy criteria, we recommend conducting empirical analyses to determine the number and types of constructs that are represented in the accountability system and how they align to the theory of action. That is, does the additional measure actually tell you something different than the other academic measures in the system if it intended to do so?

Prioritization

As noted above, practitioners should use both the policy and technical considerations when identifying and selecting measures for an accountability system. It is likely that during design and development, a measure or indicator may not support both policy and technical needs. That is, there may be a disconnect between the quality of a metric from a technical standpoint versus the desire to include a metric to promote policy or incentivize behavior. Practitioners should be mindful of possible tension between the two and determine how to support the policy goals in an accountability system. If a measure is in conflict with either the policy or technical considerations, it is important that one documents how and why it is in conflict.

For example, policy-makers may want to incentivize demonstrating problem solving or innovative thinking that might be observed in a project or other exhibition. However, the evaluation of student performance on these exhibitions is not yet sufficiently reliable and valid to support establish technical criteria for accountability system and may be vulnerable to corruption. Once these issues are articulated, steps should be taken to determine how these problems will be addressed once the system is in operation. Additionally, practitioners should consider how to minimize any effect on the overall

9 AERA, APA, & NCME, & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

[10] Hill, R. & DePascale, C. (2002). *Determining the reliability of school scores.* National Center for the Improvement of Educational Assessment. Dover, NH. http://www.nciea.org/publications/CCSSO02_Reliability_RHCD03.pdf

system by reducing its impact on the system (e.g., lowering the weight of the indicator) and documenting why that is an appropriate step for any future questions or discussion.

Dimensions that Influence Policy and Technical Considerations

The policy and technical considerations are a first step in considering what measures support the theory of action of an accountability system. The types of indicators dictated by state or federal legislation also play a large role. The following section describes how the requirements of ESSA can be approached in conjunction with using policy and technical considerations.

Considering Policy and Technical Criteria through the Requirements of ESSA

The identification and testing of system indicators should rely on the theory of action forwarded by a state to determine the overall system structure and reporting approach. For example, if a state were to focus on providing sequential indicators that support judgments of student progress toward graduation in a relatively traditional system, they may select academic indicators that are based widely on standardized achievement tests. This system might include the following:

- Year-to-year growth as an indicator of progress on the state's standards;
- Growth-to-standard as an indicator of distance to proficiency;
- Proficiency as an indicator of readiness for the next grade level content/ELP as an indicator of English language readiness; and
- Graduation rates as an overall signal of readiness.

As one would expect, the measures within each of these indicators could vary significantly depending on the types of data available, current initiatives in place at the state, and whether the measures fit well within the accountability system. That is, each indicator's measure should be quantitatively vetted to ensure its appropriateness for inclusion as a supporting the goals of the accountability system. The following section forwards some options to stoke practitioners thinking around satisfying the requirements set forth in ESSA.

Achievement

Achievement as an indicator may initially seem somewhat limited given its association with proficiency. It may be inferred from the need for differentiation in the accountability system that it would be beneficial that the measure of achievement would exhibit a high degree of variability. However, it may be likely that the range of percent proficient will be restricted to the middle of the 0%-100% range (e.g., 30-70%), depending on the stringency of the cut score. When examining how achievement is being measured, it is important to keep in mind the policy and technical considerations. What signals are being communicated? What goal is being prioritized? How consistent is the measure over time? Does there appear to be equity in the measure when compared to characteristics of students and schools? By examining alternative ways to satisfy this indicator and answering these questions (e.g., mean scale score, distance of scale score to target), evidence can be gathered in support of the accountability system. More detailed suggestions to examine achievement are provided in *Appendix Y*.

Growth or Achievement Gaps

The inclusion of a growth or achievement gap metric is again dependent on the goals stated in the state's theory of action. Further, states vary on their growth or gap measure, which will affect the metric. Given the potential differences a state may exhibit with this indicator, we recommend examining the shape and spread of the data to understand how data can be incorporated into the larger system.

Again, returning to the policy and technical considerations, practitioners should ensure that the way in which growth or gap calculations are conducted serve both the policy and technical goals of the system. Are the measures reflecting progress as intended by the selected indicators? Do they demonstrate sufficient variability to differentiate schools? Are they exhibiting enough equity that they are not unduly influenced by the composition of the students within a school or district? More detailed suggestions to examine growth or achievement gaps are provided in *Appendix Y.*

English Language Proficiency

The inclusion of English Language Proficiency poses a different kind of problem in that it may not be a ubiquitous measure for all schools—particularly in states with a homogenous population. English Language Proficiency (ELP) rates have traditionally been reported at the district level using Title III Annual Measureable Achievement Objectives (AMAOs). However, it appears that this requirement speaks to an increased focus in equity of educational access and performance.

States should begin by establishing a minimum n size for subgroup identification. Smaller n sizes typically allow as many students as possible to be taken into account for a school's performance. Another outcome may include identifying more schools with students who would take an ELP assessment. States should conduct exploratory analyses to determine the percentage of schools with subgroups, including those students who would take an ELP assessment, using various n sizes in alignment with their theory of action.

Practitioners will need to be especially attuned to issues around equity, availability, reliability, and comparability. Given the potentially different concentrations of English Learners, it may be difficult to make comparisons across schools. Further, depending on the transiency of the state population or the exit patterns of ELs, the consistency of data will be a factor to consider. States may also want to consider identifying ways to award credit to schools that are able to exit students from English Language Learner status to avoid unintended consequences associated with deflating EL proficiency rates. More detailed suggestions to consider ELP are provided in *Appendix Y.*

Graduation Rate

Graduation rates for accountability have historically relied on the 4 year adjusted cohort rate (ACR). However, accountability workbooks and later Requests for Flexibility from ESEA have also utilized the 5 and 6-year ACRs. These may be a more appropriate indicator of graduation given the shift of post-secondary governing bodies using a 4 and 6-year graduation timeline for community colleges and 4-year institutions, respectively.

In examining graduation rates, the selection of the ACR timeframe should be clearly aligned to the theory of action. That is, there should be a justification that states can identify to justify the selection. By examining longitudinal data associated with K12 and post-secondary performance (if available), practitioners may be able to accumulate evidence that speaks to the appropriateness of using a 5- or 6-year ACR, aside from the identification of comprehensive support schools. More detailed suggestions to examine graduation rate are provided in *Appendix Y.*

School Quality/Success

Due to the potential flexibility with the 5th indicator in ESSA, measures of school quality or success, a clearly stated theory of action is imperative to appropriately select measures within the state's context. For example, a state may either want to identify indicators that represent deeper measures of student learning (e.g., including local performance assessments), expand the construct of school quality beyond academic achievement (e.g., measures of climate, community engagement, or student engagement), or focus on measures of college or career readiness or access (e.g., dual enrollment, AP performance, or AP offerings).

Given the language in ESSA (and the potential lack of standardization across districts and schools), it may be prudent to assign a relatively low weight to this indicator when comparing it to the other four indicators. Additionally, overall performance targets may factor in this weight and establish targets that enable the 5th indicator to serve as points that could compensate for performance on other measures (akin to bonus points). More information on how to evaluate the 5th indicator is provided in the section titled **Considering the "5th Indicator" in ESSA Accountability.**

Hypothetical Scenarios to Showcase the Impact of ESSA Legislation

The following section of the paper presents 3 hypothetical scenarios that focus on systems under ESSA requirements with different combinations of the indicators. Additionally, it discusses the factors that influence the impact of indicators in an accountability system. For example, is the inclusion of the indicator likely to create more or less favorable ratings? Is the influence likely to be consistent for all student groups or schools? How will the distribution or spread of scores change? The next section depicts how each of these scenarios is impacted by how the characteristics of the data, the performance standard, and the aggregation method could potentially influence the accountability system. Each of these scenarios also uses summaries for a set of example schools to directly illustrate how weights and/or performance standards may impact decisions. These data will be held constant to help illustrate the impact of design decisions across the three scenarios, but the indicators may vary slightly depending on the focus of the scenarios. Readers should note that while the decisions associated with the system design are dependent on the agency's theory of action and policy needs, a strong evidence-based rationale should be established to help contribute to a validity argument of the accountability system.

Scenario 1 – Status Quo/Tight Focus on Academic Achievement

This scenario presents a system that has opted for a focused set of indicators honing in on a traditional view of school quality. This state has prioritized proficiency as the end goal, with the assumption that

proficiency is a sufficient indicator of student preparedness for college and careers. They believe that a heightened focus on academic performance toward a standard—similar to what was federally required under NCLB—will help keep conversations and efforts purposeful. To further this idea, the additional academic indicator selected is student growth with an emphasis on growth to standard (i.e., the rate at which students are mastering the state's content standards with a target of proficiency in a reasonable amount of time). Because of the desire to maintain efforts on academic performance, the state includes a safe 5th indicator, attendance, as a small portion of the system—5% of the overall rating. The example schools presented below will be used to illustrate this scenario in Table 1.

Table 1. Example School Data for Scenario 1

School	% meeting standard in ELA	% meeting standard in Math	Observed growth in ELA	Observed growth in math	% meeting growth to standard in ELA	% meeting growth to standard in math	% meeting ELP standard	Attend Rate	Grad Rate
1	45%	43%	55%	58%	55%	52%	75%	97%	88%
2	37%	38%	54%	50%	42%	39%	66%	98%	74%
3	49%	48%	68%	67%	53%	55%	70%	95%	80%
4	35%	31%	52%	44%	45%	35%	57%	96%	68%

As seen in the table above, these four schools range in performance and growth but are relatively similar with regard to attendance rates. In this scenario, the indicators are based on various sources of data. Academic performance is based on the percent of students meeting the standard in math and ELA. Observed growth is based on the percent of students who are demonstrating some policy-defined level of growth in math and ELA (e.g., at least typical growth using student growth percentiles, SGP). Growth to standard is based on the percent of students who are on track to be proficient in 3 years. Alternatively growth could also be conceptualized using an educator value-added assessment system (EVAAS) for observed growth (i.e., using ranges of positive or negative school values to show the relative gain or loss across years) and growth to standard (i.e., using projections to compare actual vs. expected performance).

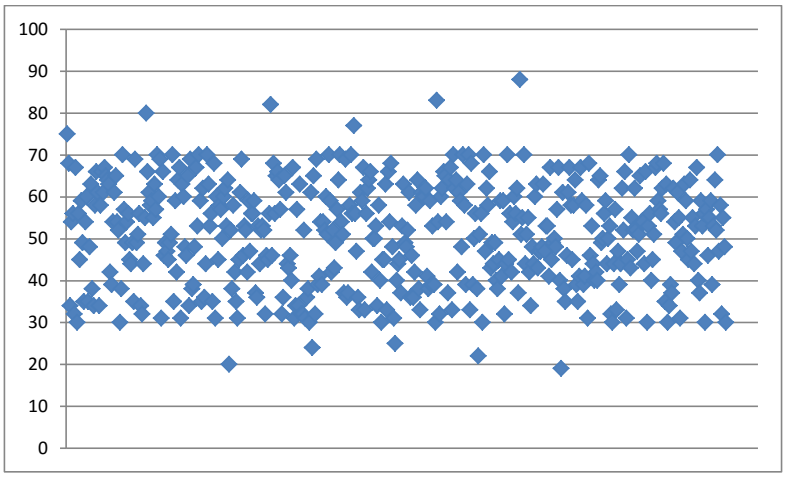
If one were to evaluate how these data are applied under an accountability model, there are two key aspects of the system that should be addressed—the performance standards within each indicator and the aggregation method across all of the indicators. These two concepts are discussed in greater detail below.

Performance standards. The performance standard for any given indicator has implications for the accountability system as a whole. While not an exhaustive list, the performance standard influences (1) what signals the indicator communicates, (2) how well the indicator differentiates in isolation, and (3) how the indicator contributes to the variability of the entire accountability system. This has historically been evident when states adopt new cut scores for assessments. As part of the transition to college- and career-ready assessments, the increased rigor in cut scores has typically yielded lower rates of proficiency. As a result, the percent of students meeting the standard aggregated at the school level

tends to have a more restricted range for the majority of schools in a state. This is illustrated in the example data below.

EXAMPLE BELOW (currently placeholder)

Comment [JD2]: Real data would be good to include just to demonstrate that we really aren't working with the entire range. Although, a figure may not provide any value at all.



In addition to the performance standard on the assessment affecting variability, how points are awarded in the system (if for example an index is used) will also determine how the accountability system identifies schools.

This same issue exists for conceptualizations of growth. Regardless of what growth model a state uses, SEAs will need to determine the policy decisions that determine what level of growth is sufficient to reflect that a school is making adequate progress. With SGPs, for example, many states have opted to use typical or better as a sufficient criterion for observed growth (out of the options low, typical, and high growth). However, one should examine the actual school-level distributions, determine whether additional growth levels are necessary to maximize differentiation, and how those policy decisions affect influence communications. Practitioners will then need to determine how each of the growth levels (or ranges) would contribute to a school's score or decision rules.

Aggregation method. As states consider the various indicators that best align with their theories of action, agency leaders must also consider the ability of the system to meaningful differentiate schools on an annual basis. These comprehensive summative ratings require a minimum of 3 performance levels, where the highest and lowest rated schools cannot be in the same category. Further, the first four indicators must have "much greater" weight than the 5th school quality/student success indicator. Ultimately, the way in which the indicators are aggregated must support the intended use of the accountability system. Does the system prioritize high-stakes application, providing performance information, or informing continuous improvement in instruction and school improvement efforts? Clearly defining the purpose will help clarify how the aggregation method should be calculated.

Using the example in the first scenario above, there are many ways one may want to define the weights of each of the indicators. While policy-defined weights help reflect the values of the system or agency, they should be considered carefully. For example, weighting proficiency most heavily, followed by growth to standard, and observed growth as the slightest indicator communicates a priority of getting students to proficiency. However, if school-aggregate growth data exhibit a higher degree of variability when compared to proficiency, growth will effectively drive the differentiation of schools. It is for this reason that it may make good sense to weight non-differentiating indicators lightly. For example, weighting attendance as only 5% of the whole system as it doesn't truly differentiate (i.e., almost all schools will be awarded all possible points on attendance). Under this aggregation example, it is likely that higher performing schools will perform more favorably in the system regardless of their performance on growth. However, allowing observed growth to contribute more heavily than proficiency will also change the makeup of high performing schools. That is, higher growth in schools can compensate for lower performance as a means to highlight different profiles of performance.

Alternatively, should a state prefer to use a decision matrix, it might require a determination of progress. That is, on how many indicators has a school made progress this year? This assumes that the accountability system's indicators have individual targets that have been vetted by a design committee. For elementary and middle schools (i.e., a total of 8 indicators excluding graduation rates above), a school's rating may be dependent on whether sufficient progress has been demonstrated on a certain number of indicators. For example, ratings may be determined by progress on less than 3 indicators, 3-4 indicators, 5-6 indicators, or 7 or more indicators yielding different ratings. While this is a simplified example, a design committee would have to determine what constraints they would want to place on the accountability system to ensure that school determination reflect system values.

Aside from the performance standard within the indicator, SEAs will need to consider what performance targets are reasonable for each indicator or the accountability system as a whole. An immediate and widely experienced challenge associated with this is the case of unrealistic long-term targets (e.g., 100% proficient by 2014). The idea of ambitious but reasonable goals (see Linn, 2003[11]) is not new and ESSA affords agencies an opportunity to think carefully about the expectations developed for schools. By establishing targets that are already being met in a proportion of schools (e.g., 75th percentile for all schools by 2025), it is more likely that school improvement efforts will be taken more seriously and that more schools will demonstrate improvement toward the target.

Comment [JD3]: Please check my thinking, but I'm reading that there will need to be defined targets not just for the overall "score," but also for the individual indicators?

And should this be reflected in the sections below? I'm thinking it's stated once and that is likely enough.

Regardless of the target defined for each indicator or for the accountability system, SEAs and their planning committees should at least examine historical data to determine how difficult it may be to close the distance to the target. The trend data on any of the previous indicators can help inform the degree to which adjustments will be required to maintain or accelerate improvement over time.

Scenario 2 – Closing Achievement Gaps

[11] Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Center for the Study of Evaluation. National Center for Research on Evaluation, Standards, and Student Testing:* Los Angeles, CA.

This scenario presents a system that prioritizes the concept of equity in access to educational opportunity in a diverse state. As a result, this state believes that a focus on the difference between subgroup and non-subgroup performance should be a primary determinant for designing programs and allocating resources. Furthermore, this state believes that to appropriately address subgroup needs, there must be a consideration of non-academic school quality conditions. Thus, this state has defined the following 5 indicators: academic proficiency, achievement gaps by mean scale score, graduation rate for high schools, English language proficiency, and a measure of school climate. This state has also set an n-size of 15 for subgroups to ensure schools are able to identify gaps where they exist.

Table 2. Example School Data for Scenario 2

School	% meeting standard in ELA	% meeting standard in Math	% of gap closed in ELA	% of gap closed in math	% meeting ELP standard	School climate favorability rating	Grad Rate
1	45%	43%	55%	60%	75%	85%	88%
2	37%	38%	85%	75%	66%	90%	74%
3	49%	48%	78%	67%	70%	82%	80%
4	35%	31%	80%	92%	57%	93%	68%

As seen in the table above, these four schools demonstrate various differences depending on the indicator. For example, there is relative similarity in the school climate favorability rating (e.g., percent of respondents who generally agree or strongly agree the school provides a climate favorable for learning) but there is a wider difference in how close schools are to closing their achievement gaps (e.g., 100% of the gap is closed) and some difference in the % of students meeting the standard in math or ELA. Again, the way in which performance standards and aggregations are defined can have major implications on the designations of schools.

Performance standards. As noted previously, the performance standard influences (1) what signals the indicator communicates, (2) how well the indicator differentiates in isolation, and (3) how the indicator contributes to the variability of the entire accountability system. In the example above, two indicators will likely be under the greatest amount of scrutiny: achievement gap and school climate. Achievement gaps can be calculated in a number of ways that seek to balance accuracy and simplicity. While it can be argued that simply subtracting percent proficient between the subgroup and all students is an easy calculation approach, it does not take into account the nuance of membership and meaningfulness of the metric.

For example, there is a high degree of likelihood that historically underserved racial groups will also be present in the low-socioeconomic subgroup and English Language Learner subgroup. Therefore, it may be more appropriate to conceptualize the achievement gap as the difference between the top and bottom performers (e.g., top 25th and bottom 25th percentiles). Furthermore how the achievement gap is calculated is important. In this case, the achievement gap is determined by the following:

$$\text{Achievement gap closure} = \frac{\text{Mean Scale Score of the bottom } 25^{th} \text{ percentile}}{\text{Mean Scale Score of the top } 25^{th} \text{ percentile}}$$

This example allows for a natural awarding of points with cases of schools with no gaps being awarded the full amount of points. Another example may be using the difference in mean scale scores that are then converted to a common scale to aggregate across grades and domains. An alternative may be to simply subtract the two and determine the total shortfall of the lowest performing students. This number could then be compared statewide and anything beyond a certain amount (e.g., +/- 1 standard deviation) would result in fewer points (e.g., more than 1 standard deviation below the average) or in bonus points (e.g., more than 1 standard deviation above the average). Regardless of the approach taken to calculate the achievement gap, historical data and comparisons to other information should be made to ensure there is sufficient variability in the data and that it does not duplicate some unrelated measure (e.g., low socio-economic status rates).

Similarly, this would need to be taken into account for the climate measure. Because there is less historical information on climate or school quality measures, it may be especially beneficial to understand the spread, skew, and general shape of the data. Further, understanding what kinds of relationships may exist with other indicators will help inform how performance standards are defined. Implications of using less known indicators as part of school quality/student success measures are discussed in the 5th indicator section below.

Aggregation Method. As previously noted, the ways in which indicators may be aggregated should be informed in large part by the theory of action. However, the theory of action should be considered a starting point for policy-grounded weights and depending on how the data behave, can be adjusted using empirical evidence.

Using the example in the second scenario above, the indicators that are given the most weight will communicate the biggest priority for the state. A relevant risk with achievement gaps can be seen in school 4. The relatively low performance exhibited in both ELA and math may be masking issues with performance in the school. That is, the lack of achievement gap is likely an artifact of an underperforming top 25 percent rather than a small difference between the bottom and top 25 percent. In this case, regardless of the performance standard applied, school 4 would not likely be identified as having an achievement gap. Thus, a more significantly weighted achievement gap indicator would result in this school being rated highly, but that may not accurately reflect the school's performance across all indicators (given the lower performance of all students in ELA and math and the lower EL proficiency rate). For this reason, it is critical that the weights support the theory of action being forwarded in the accountability system.

When considering the impact of the school climate indicator, again school 4 has a relatively higher rating than the other schools in the table above. This may be a product of the school effectively engaging parents and establishing a climate of positivity as rated by students and teachers. This climate, however, may not yet be resulting in an impact on learning outcomes. Alternatively, the conditions and demographics of the school (e.g., an urban school with high truancy rates) may result in more favorable ratings due to students feeling engaged regardless of whether any objective measure of engagement is present. Again, depending on the principles the system wishes to prioritize and the behaviors that the state wants to incentivize, this may be an appropriate model to apply.

Given the example presented above, the theory of action and subsequent outcomes must be examined for congruency. We recommend that SEAs examine evidence that confirm the outcomes of the accountability system (e.g., school designations, indicator distributions, disaggregated indicators) and support the intended theory of action. States should be mindful of specific use cases that may break down the assumption associated with weighting any indicator more than another. That is not to say that the same issue may not emerge if all indicators are weighted the same, for example. It is incumbent upon the agency to then ensure that the system and the signals it is trying to communicate are sufficiently tolerant of any false positives or negatives.

Scenario 3 – Measures of College- and Career-Readiness

This scenario presents a system that privileges college- and career-readiness for students exiting high schools. This state believes that their current assessment is an appropriate reflection of the knowledge, skills, and abilities necessary for students to be successful in college and careers. However, the state believes that including indicators that speak directly to post-secondary readiness are critical to ensuring the state's vision is reflected. In addition to the required end-of-year assessment, the state has decided to include the a combination of post-secondary measures that include results on career-focused measures (e.g., ACT WorkKeys, other occupational exams, completion of career-technical education courses) and college-focused measures (e.g., AP/IB performance, ACT performance, and dual or AP course results). In addition, this state is using proficiency on statewide assessments, measures of growth, English language proficiency, and graduation rate for high schools.

Table 3. Example School Data for Scenario 3

School	% meeting standard in ELA	% meeting standard in Math	Observed growth in ELA	Observed growth in math	% meeting ELP standard	% meeting college-ready indicators	% meeting career-ready indicators	% meeting CCR indicator	Grad Rate
1	45%	43%	55%	58%	75%	50%	57%	87%	88%
2	37%	38%	54%	50%	66%	35%	40%	68%	74%
3	49%	48%	68%	67%	70%	44%	34%	70%	80%
4	35%	31%	52%	44%	57%	25%	32%	48%	68%

As seen in the table above, these four schools demonstrate various differences depending on the indicator. For example, there appears to be enough diversity in the percent of students meeting college-ready indicators but more relative similarity when examining career-ready and observed growth indicators. As noted previously, the way in which performance standards and aggregations are defined can have major implications on the designations of schools.

Performance standards. As noted previously, the performance standard influences (1) what signals the indicator communicates, (2) how well the indicator differentiates in isolation, and (3) how the indicator contributes to the variability of the entire accountability system. In the example above, the two indicators of college- and career-ready indicators are the most relevant for discussion. Because this state believes that the paths for students somewhat differ between college and career, the agency has opted to distinguish measures of college-readiness and career-readiness. Therefore, the accountability system

will report on college- and career-ready measures separately, but will take a disjunctive approach for a high-stakes calculation of whether students are college- **or** career-ready (i.e., % of students meeting the CCR indicator).

Conversely, other states may believe that it is more appropriate to use a more conjunctive approach in determining college- and career-readiness. For example, other states may believe that a test-based measure is a sufficient indicator for CCR, but LEAs may be required to provide additional evidence for less objective indicators. Thus, LEAs may be required to select one of the available performance measures (i.e., tests or exams) in addition to completion/enrollment in career- and college-focused courses.

It will take careful consideration of what indicators are available and defensible to ensure there is not an uneven playing field for those schools and districts with fewer resources or opportunities for students. Whether states choose separate college-ready measures (e.g., state summative assessment, ACT/SAT score, AP, IB, or dual-ready courses), career-ready measures in conjunction with a defined assessment (e.g., pre-defined score on state assessment), or career-ready measures as a standalone indicator of readiness (e.g., workplace experience courses, completion of a co-op job training opportunity, ACT WorkKeys), performance targets will need to be considered within the accountability system.

If one were to apply the disjunctive model as described above, the calculation might be relatively simple:

$$\frac{\text{\# of students who meet either the college- or career-ready criteria}}{\text{Total \# of students enrolled}}$$

With any kind of calculation, we recommend at least 2 pieces of data to examine how performance targets might apply within the accountability system. The first set of data will likely be the easiest to obtain, but both can be useful. These include the following:

1. Historical trend data on any of the previous indicators to inform the degree to which adjustments will be required to maintain or accelerate improvement over time; and
2. Any available data that links high-school performance on each of the indicators to post-secondary data. This connection can help validate whether proxies in high school reflect success in college or careers.

Once performance standards and targets are determined, agencies can also consider how aggregation rules are applied to this indicator.

Aggregation Method. As previously noted, the ways in which indicators may be aggregated should be informed in large part by the theory of action. However, the theory of action should be considered a starting point for policy-grounded weights and depending on how the data behave, can be adjusted using empirical evidence. It is important to note that these potential indicators of college- and career-readiness are proxy measures. Thus, the relative weight of this indicator may be best informed by the amount of validating data that an agency has available to establish evidence-based claims around

readiness. If little data exists, it may be more appropriate to make this indicator have less of an impact on the overall rating of schools. However, if states have connected their K-12 and post-secondary data systems, it may be easier to verify claims of readiness and justify more impact on overall school ratings.

Additionally, states will need to consider whether they are going to treat process and outcome measures equally. For example, does enrollment in a dual credit class provide the same inference as a minimum grade in a dual enrollment course? Should the accountability system incentivize the provision of college- and career-ready focused courses? How universally available are these courses across the state? Depending on the answer to these and other questions, accountability teams will clarify their theories of action to determine what evidence corresponds to the hypothesized mechanisms and processes.

Considering the "5th Indicator" is ESSA Accountability[12]

Most of the indicators required under ESSA are at least familiar, even if the specific metrics proposed may be new under ESSA. However, the types of metrics and indicators suggested for the "fifth indicator" are relatively new and generally have not been used in accountability systems. There are several psychometric characteristics required of this indicator—valid, reliable, and must differentiate performance—but, in general, the options for what can be used as an indicator are fairly wide open. That being said, it will be important to consider each of these technical requirements as one thinks about potential indicators. While reliability is easily defined, the validity of an indicator (within a system context) is less clear but needs to be based on a well-articulated theory of action. Current thinking about "differentiate" is that the law intends for indicators to have a fair amount of true variability among schools compared with indicators such as elementary school attendance that essentially acts as a constant in the system.

Conceptualizing the Other Measure of School Quality

States need to be thoughtful about this additional indicator regarding how it fits with their conceptions of educational accountability and school quality. Do state leaders think this additional indicator will 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 precursors to the achievement and growth academic indicators? For example, some might want to include an indicator of student engagement because they think it is a precursor to higher levels of student achievement, while certain social-emotional learning indicators help broaden our characterizations of school quality. Obviously, there is considerable overlap among these conceptions. We expand these conceptualizations below. We argue that this indicator can be characterized along the following dimensions:

- ✓ School or student unit of analysis
- ✓ Level of inference

[12] From Marion, S. & Lyons, S. (2016). In search of unicorns: Considering the "5th Indicator" in ESSA accountability. *National Center for the Improvement of Educational Assessment:* Dover, NH.

- ✓ Potential corruptibility
- ✓ Level of data burden

School quality or student success. This is likely the most important dimension on which to conceptualize the fifth indicator. School quality indicators are intended to broaden the ways in which we characterize school effectiveness beyond the typical indicators of reading and mathematics achievement. Generally, the data are collected through individual students, but the school is the unit of analysis. On the other hand, data for student success indicators are collected and reported at the student level.

Level of inference. The level of inference associated with the indicator is an important dimension and one that interacts with the next dimension we discuss; potential corruptibility. A high inference indicator might be something like school climate where 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. As one can see, there are multiple steps along this inferential chain that must be validated to substantiate the claim that one is indeed measuring school climate. Student engagement might be a student success indicator that falls along the higher inference end of the continuum. Even if it is operationalized as something like attendance, strong inferences are required to support the notion that attendance is an accurate indicator of true 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 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.

Potential corruptibility. 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 through surveys or other similar approaches. We must carefully consider "Campbell's Law" when using any indicator, but especially those that are easily corruptible if they are used as part of a high stakes (or at least publicly reported) accountability systems.

> *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*[13].

While over 40 years old, this "law" has been well documented far too often since 1976, but 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 really thoughtful and careful in our accountability system design and especially in the design of this fifth indicator.

[13] Campbell, Donald T., Assessing the Impact of Planned Social Change. The Public Affairs Center, Dartmouth College, Hanover New Hampshire, USA. December, 1976.

One of the ways we may minimize the 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 3-4 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 that required clear demonstrations of evidence where corruption may be minimized such as some of the lower inference indicators discussed above.

Level of 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 low on the data burden end of the 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 lower-skilled workers raising some data quality concerns. We are not necessarily advocating for considering only measures with low data burden, but we argue that such burden needs to be considered carefully.

Validation and Theories of Action

The considerations and dimensions discussed above are important, because it highlights how one approaches the development and validation of the 5th indicator. 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. For example, one might think of a school engagement initiative that encourages students to participate in community service or other applied projects. Such engagement may be thought to help students hone leadership skills and other characteristics associated with being responsible global citizens, which are not measured well on tests. It stands to reason, then, that validating the indicator with assessment data would be misplaced. Rather, we would seek other data thought to affirm our understanding of the construct. On the other hand, one might operate from a perspective 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 is less certain. Importantly, these conceptions are both potential hypotheses that should be clearly articulated in a theory of action and then evaluated empirically as data are collected.

As one can see, we recommend a careful examination of the 5th indicator, as we have recommended for any indicator in an accountability system. These indicators should be reflective of the theory of action and practitioners should collect evidence where possible to support the claims being made in the theory of action. For a full discussion of the 5th indicator, please see Appendix Z.

Conclusion

The examination of indicators and how they contribute to an overall score should be a recursive process. ESSA exhibits some similarity to the Requests for Flexibility from ESEA, with greater flexibility in how supports and rewards are provided to schools and districts. Furthermore, states may already have well-developed a theories of action, indicators, and associated weights. However, the extent to which simulations have also included inferential analyses likely varies significantly. These steps could be used to reinforce the selection of indicators and defined weights within a state's theory of action.

Policy and technical criteria. By applying the policy and technical criteria to already developed systems, states have an opportunity to revisit assumptions and gather evidence in support of a validity argument for their accountability systems. For those states that are revising or revisiting their accountability system design, the technical and policy criteria serve as example filters that can encourage the inclusion of additional criteria most relevant to each agency's context.

Characteristics of data. Practitioners should also evaluate the impact of data characteristics on an accountability system. A sound accountability system is dependent in part on minimizing disparate or inequitable outcomes based on the performance of students/schools on indicators. For example, how do measures of interest interact with the characteristics of the population to which they will be applied? Do certain indicators exhibit behaviors or patterns when applied to smaller vs. larger schools? How practically available are the measures or how do they behave over time? Does the presence of subgroups affect the outcomes of the indicators and/or the aggregation of indicators? How do school types or grades influence how schools are differentiated? By attending to data characteristics in tests and simulations, unintended consequences may be identified and avoided.

Performance standards. In addition to the various dimensions and criteria presented, practitioners should evaluate the impact of performance standards on the accountability system. Outcomes can be influenced greatly depending on how one sets a performance threshold for each indicator. When establishing performance standards, there are many approaches that a state can take. For example, one can use norm referenced approaches (e.g. divide the distribution into quartiles and award points for each of the 4 categories), criterion referenced approaches (e.g. award points based on an established benchmark like an AP score of 3 or higher), judgment based approaches (e.g. convene a panel and have them set thresholds), or a combination. By considering the indicator and the performance standard of the indicator, practitioners can balance the two and support a valid accountability system. That is, a 'rigorous' indicator can be made to have a more favorable impact by virtue of the established performance standard or vice versa. Considering realistic, yet rigorous performance standards will help contribute to an accountability system that produces valid results and incentivizes desired behaviors.

Aggregating indicators. While the selection of indicators is contextualized to a theory of action, considering how these indicators interact may have a more generalized set of steps. That is, using simulated data, one can approximate the impact on how each set of indicators and their relative weights differentiate schools. There are classes of analyses that may help policy makers better understanding the unique contributions of indicators and their influence in accountability models. However, because

there is no “best number” that differentiates school quality, the relative weights of the accountability system should be selected theoretically—in this case, by using the state’s theory of action. The initial weights and indicators can then be checked analytically to maximize differentiation using the stated theory of action. For suggested approaches, please see *Appendix Y.*

The aggregating of indicators becomes an exercise in balance. Practitioners may believe that a system that prioritizes the differentiation of schools and does so with complicated weighting structures is more accurate. However, it may be more difficult to explain when garnering buy-in and could be at risk of total rejection. As noted in the earlier design and subgroup section, choosing a conjunctive, compensatory, or hybrid aggregation method can have effects on the number and types of schools identified with a certain score or in each category. Alternative classification approaches, such as using a decision matrix of indicator performance, will also have their own classification tendencies. It is imperative that practitioners examine how the aggregation results behave to promote an equitable operationalization of the theory of action. Including this examination can help practitioners gather evidence that the accountability system is accurate and reflective of the stated goals.

We hope that this paper offers tangible and practical suggestions for those designing or designing accountability systems under ESSA.

Appendix X

Comment [JD4]: These appendices are intended to be deeper dives into the sections above that reference the relevant appendix section.

Educational accountability systems should be designed to contribute to the improvement of educational systems[14]. However, there is also a need to define how the accountability system defines the quality of education and how it is communicated to stakeholders (e.g., educators, administrators, policy-makers, community members, and students). There are two primary sets of factors that should minimally be considered when evaluating the efficacy of accountability systems. These include (1) policy considerations and (2) technical considerations.

While both are important guiding factors, it is possible that there may be a tension between the two. That is, an indicator may have a high policy value but may not perform well against technical criteria for inclusion in an accountability system. In this case, the theory of action should be used as a guide to determine whether the measure should be a part of the accountability system. The remainder of this section forwards a sample set of considerations and should not be taken as an exhaustive list. Rather, other criteria should also be included depending on the system and its components that are deemed important by the relevant stakeholders and practitioners.

Policy goals. The overall policy goals set forth in the accountability system should be a primary driver in designing and evaluating an accountability system. For example, the two goals of (1) increasing the graduation rate of students and (2) increasing the rates of college and career readiness (CCR) for all students communicate different expectations. Furthermore, these two goals also highlight the need for potentially different measures.

Should a state include both goals in their accountability system, policy makers would need to consider measures that reflect progress against both. More specifically, the graduation rate itself may be a necessary but not sufficient indicator of CCR. In fact, it might be considered a more proximal indicator of overall CCR. Further, it may be beneficial to include additional proximal (and potentially non-high stakes) indicators that provide insight toward hitting graduation rate targets (e.g., credit completion flags for students once they hit grade 9, absenteeism, etc.). These additional measures could be included in reporting but not as part of high-stakes calculations. If there isn't a clear connection between proficiency or credit accumulation and CCR, additional work may be necessary to establish/select a viable metric that communicates this signal.

Equity. Equity in accountability systems seek to incentivize actions that lead to academic improvement for the lowest performing students and in detecting desired outcomes (see Domaleski & Perie, 2012[15] for more detail on approaches to emphasize equity in accountability systems). Equity in education also comprises fairness (i.e., characteristics of the student or school do not interfere with demonstrating educational ability) and inclusion (i.e., the ability to demonstrate educational abilities is universally

[14] Hargreaves, A. & Braun, H. (2013). *Data-Driven Improvement and Accountability.* Boulder, CO; National Education Policy Center. Retrieved April 13, 2016 from http://nepc.colorado.edu/publication/data-driven-improvement-accountability/.
[15] Domaleski, C. & Perie, M. (2012). *Promoting equity in state education accountability systems.* National Center for the Improvement of Educational Assessment. Dover, NH. http://www.nciea.org/publication_PDFs/ Promoting%20Equity%20CSDMP110712.pdf

accessible)[16]. These dual aspects of equity can help filter the measures that one may select for inclusion in an accountability system. That is, how well do the measures identified uphold the ideas of fairness and inclusion? For example, does the measure selected seem to describe students and schools fairly while not being adversely affected by student or school characteristics? While we recommend that this is more thoroughly investigated (please see the technical considerations section below for more detail), answering this question at face value early can help avoid unintended consequences. Additionally, does the measure selected include all students and schools? Any indicator in an accountability system should provide all schools the opportunity to demonstrate their students' progress.

Promoting college- and career-readiness. Many different kinds of goals can be proposed during accountability design and development. In this case, we will focus on one that promotes college- and career-readiness. This kind of goal will require practitioners to consider what components and processes in the theory of action contribute to summative (i.e., end-of-process) indicators of CCR. By doing so, one can potentially distinguish between potentially high-stakes and low-stakes indicators. That is, the long-term indicator of CCR that may be more appropriate to include as part of a high-stakes differentiation measure while the short-term and mid-term indicators may be restricted to low-stakes use. For example, the short- and mid-term indicators may be best used as part of low-stakes reporting that informs overall progress.

Focusing on other measures of school quality. A potentially powerful aspect of ESSA is the inclusion of additional measures of school quality. One can conceptualize this as a means to broaden the construct of school quality or as a way to increase the depth of current measures of school quality. These ideas are described in further detail below.

*Broadening the construct of school quality***.** The idea of broadening the construct of school quality includes going beyond traditional academic indicators. Essentially, this widens the view of school performance to include those indicators that could paint a more holistic picture of school quality. For example, some states have included measures of school climate, program reviews of offerings, parent engagement, and participation in physical education programs as indicated in their ESEA Flexibility Requests. While there is potentially unique value that these types of measures could add to understanding school quality, they may also be more susceptible to corruptibility or non-standardized data collection making comparisons difficult.

*Increasing the depth of current academic measures of school quality***.** An alternative approach to measuring school quality would be to maintain a focus on academic success but seeking deeper measures of that success. This widens the view of school quality but in a different manner than encompassing a more holistic view. Rather, it may serve to highlight certain mechanisms that support academic success. For example, this kind of focus may privilege capturing higher order thinking skills or understanding the application of content more thoroughly. This might include standards-based reporting at the local level, performance based assessments, portfolios, or enhanced assessments that

[16] OECD (2008). *Ten Steps to Equity in Education.* Retrieved April 22, 2018 from http://www.oecd.org/edu/school/39989494.pdf

target process and synthesis. Further, it may be possible that the selection of measures can also provide insight into precursor skills or more proximal indicators of school quality.

Resistance to corruptibility. A final example technical criterion to consider is the degree to which a measure is resistant to corruptibility. There are a few dimensions that may influence how resistant a measure may be to corruption. These could include process/outcome measures, objective/subjective data, and high-stakes/low-stakes use. With regard to these three dimensions, each serves a different purpose where the theory of action should inform they type of measure that is used. For example, a process or outcome measure (i.e., enrollment in a course vs. end-of-course grades) both reflect different elements of school quality (i.e., access to education vs. performance) and differing resistance to corruption. The former is more easily corrupted than the latter, especially if performance is gauged on a standardized test.

When considering the objectivity or subjectivity of a measure, the reader is reminded of the previous section on policy goals and the intent that the accountability system should incentivize certain types of behaviors. It may be possible that subjective measures (e.g., socio-emotional measures, school climate indicators, instructional engagement inventories) may support this in novel ways that objective measures may not have accomplished in the past. However, the more subjective a measure, the more easily it can be corrupted. If a high-stakes measure is a more appropriate selection, one might consider supplementing it with more low-stakes information, which may take the form of a process or subjective indicator.

Even objective, summative, high-stakes indicators can be subject to corruptibility. For example, growth scores can be artificially inflated by reducing performance on prior score points. While this is perhaps more difficult to do in practice (and egregious in nature), no measure is immune to corruption. However, balancing the purpose of the measure and its corruptibility is an important factor in selecting a measure.

Utility. As accountability systems are put into place, states must examine how well the system supports the intended goals without deteriorating other aspects of the educational system. Does the accountability system serve the practical information and behavioral needs of users? How timely is the feedback? Depending on the intended uses of the accountability system (overall ratings vs. informing continuous improvement), timeliness may be of great or of little importance. What measurement properties are present in the data informing the indicator? And do those measurement properties impede or contribute to the end-users understanding of how to improve on that indicator? What related supports will need to be put in place to mitigate any complexities that are associated with the measurement? The decisions that are intended to be a product of the system (e.g., personnel decisions, resource allocation, professional development selection) should then inform the types of high-stakes and related low-stakes indicators that are included for school ratings and additional reporting.

Technical Considerations

While policy considerations often serve as the initial screen for indicator selection, they should not be used in isolation. The policy considerations should be supported by appropriate technical considerations

to ensure the measures selected function in a manner that provides trustworthy and meaningful information to support the system goals. Technical considerations can be thought of as filters to ensure data behave appropriately as individual indicators and as a combined set within a system, but should uphold the policy goals of the system. These considerations include concepts such as technical goals, availability, reliability, comparability, and equity.

Technical goals. In the same way that the theory of action and the accountability system forward a set of policy goals, they should also include technical goals selected to fit the policy objectives. For example, is the goal of the accountability system to maximize differentiation on the combined measure of school quality? Or is the goal to have a measure that is as consistent as possible over time? While certain criteria will be invariant regardless of the goal (e.g., availability of the measure, robustness to corruptibility, equity), the data characteristics may need to be different depending on the technical goal of the system. Some recommended analyses are presented in Appendix Y: *Dimensions that Influence Policy and Technical Considerations*.

Availability and comparability. While seemingly distinct, the ideas of availability and comparability are interrelated. Depending on the availability of a measure or metric, there is a direct impact on the comparability among sites using different measures. This raises the question of whether every school needs to have the same data or tool. By using an indicator that has universal availability, it becomes easier to make comparisons. However, it may be possible to address differences in availability by having access to supplementary measure. That supplementary measure, however, must speak to the same types of claims made across potentially different measures. This interpretation privileges comparability as a function of the claim being made, rather than comparability as a function of the process used to make that claim.

For example, if one of the policy goals were to better understand students' ability to master grade-level content while demonstrating evidence of higher-order thinking in a given subject, a performance-based assessment may be a very appropriate example of evidence. However, a state may not have enough performance-based assessments, a process to support quality control of locally developed assessments, or not all localities may have the resources to implement a performance-based assessment. This would require the state to develop some other set of criteria that allows for students to demonstrate mastery of that content (e.g., alternative tests or locally-defined portfolios). However, this may still require some subset of information to compare across groups. So while it may be easier to ensure that comparability can be achieved by ensuring everyone has equal access to demonstrating success on the indicator, there may be avenues one can take to allow for flexibility at a local level given the resources and political will present in a state.

Reliability. Reliability is a common construct that is regularly considered in assessment and research. According to the *Standards for Educational and Psychological Measurement*[17], reliability can be considered generally as the consistency across replications of a certain event as quantified using a variety of metrics (e.g., reliability indices, measurement error, sampling error, etc.). However, readers

17 AERA, APA, & NCME, & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

are encouraged to review Hill and DePascale's (2002[18]) discussion on methods to determine the reliability of an accountability system. In state's assessment programs, examinations of reliability include—but are not limited to—analyses of internal consistency, error of measurement, classification accuracy, and classification consistency. When considering measures in an accountability system, reliability may include two areas of focus. First, it would include considerations of the consistency of the metric. Second, it would include considerations of the consistency of the included population.

With regard to the metric, what are the upstream factors that may influence the downstream consistency? For example, these factors might include consistency of the content and standards across time and grades, equating considerations, motivation, administration factors, and scoring changes. With regard to the population, what characteristics of the population should be considered that may influence consistency of outcomes? For example, how transient is the population? How inclusive of students is the measure across the state? What types of idiosyncrasies are demonstrated by the population? These are some of the kinds of questions that should be answered empirically once a measure has been identified.

Equity. Equity was initially introduced as policy criteria, but should also be examined as a facet of technical criteria. One might consider the policy criteria to serve as an initial examination akin to face validity. Including equity as a technical criterion, however, serves to empirically determine whether the ideas of fairness and inclusion are upheld. One way to conceptualize equity would be to determine the relationship between the measures in an accountability system and characteristics of the population in the state. For example, what is the correlation between indicators and measures of low socio-economic status? One could interpret a lower correlation as a more equitable system than if a stronger correlation were detected. Another example may include the correlation between indicators and school size, number of subgroups detected, or percent minority of the school. While correlations are often easy to conduct and interpret, other analyses that provide greater insight into the impact of school and student characteristics on measures can be applied (e.g., predicted impact of indicators based on school characteristics, cluster analysis to determine characteristics of defined groups against indicator outcome, etc.).

Focusing on other measures of school quality. Measures of schools quality was also introduced as policy criteria. However, practitioners should examine empirically how the measures interact and within an accountability system and how they contribute to the overall school determination, classification, or score. This can be done by conducting exploratory analyses to determine how measures correlate with each other and the outcome. This can help states understand whether alternative measures of school quality are being included in the accountability system. Predictive analyses using historical data can also help determine how the individual indicators contribute to the overall score. By including a set of empirical analyses based on the system, states can collect additional evidence in support of their accountability systems.

[18] Hill, R. & DePascale, C. (2002). *Determining the reliability of school scores.* National Center for the Improvement of Educational Assessment. Dover, NH. http://www.nciea.org/publications/CCSSO02_Reliability_RHCD03.pdf

Appendix Y

Dimensions that Influence Policy and Technical Considerations

The policy and technical considerations are a first step in considering what measures support the theory of action of an accountability system. The types of indicators dictated by state or federal legislation also play a large role. The following section describes how the requirements of ESSA can be approached in conjunction with using policy and technical considerations.

Achievement

Achievement as an indicator may initially seem somewhat limited given its association with proficiency. It may be inferred from the need for differentiation in the accountability system that it would be beneficial that the measure of achievement would exhibit a high degree of variability. However, it may be highly likely that the range of percent proficient will be restricted to the middle of the 0%-100% range (e.g., 30-70%), depending on the stringency of the cut score. Suggestions to examine achievement are provided in *Appendix X.* In order to best determine how to use a proficiency indicator, the following steps should be employed:

1. Examine the distribution of proficiency rates to determine the degree of variability across the state.
2. Calculate alternative methods to using proficiency as a portion of the accountability system and compare the variability in each data set.
    a. Assign points to students at different achievement levels and multiply by the ratio of students at each performance level. Target could be 100% * 3 (e.g., Proficient level), or 75% * 3 (proficient level).
    b. For each grade within a state, calculate the distance a student's scale score is from the proficient cut.
    c. Mean scale score
    d. It is still worth, however, examining how straight proficiency would work as an amalgam of both ELA and Math. This would create a built in compensatory indicator.
3. Determine whether alternative measures provide a better predictor of the differentiated outcome than percent proficient.

When taking these steps, it is important to keep in mind the policy and technical considerations. What signals are being communicated? What goal is being prioritized? How consistent is the measure over time? Does there appear to be equity in the measure when compared to characteristics of students and schools? By examining alternative ways to satisfy this indicator and answering these questions, evidence can be gathered in support of the accountability system.

Growth or Achievement Gaps

The inclusion of a growth or achievement gap metric is again dependent on the goals stated in the state's theory of action. Further, states vary on their growth or gap measure, which will affect the

metric. Given the potential differences a state may exhibit with this indicator, the following should likely be employed to better understand how the measure itself behaves and the way in can fit most appropriately into the accountability system.

1. Utilize measures of dispersion to determine what idiosyncrasies exist in the data to avoid potential unintended consequences of seemingly logical point assignments. That is, do existing cuts minimize variability by compressing the distribution?
2. Examine alternatives to ensure aggregates of growth or gaps exhibit sufficient variability at the school levels (e.g., assigning points using stanine cut points around schoolwide observed growth or schoolwide value-added estimates; examining statewide distributions of necessary growth-to-standard estimates to assign points based on approximate "adequate" observed growth amounts).
3. Specific to achievement gaps, it may be worth considering alternatives to subtracting proficiencies from subgroups and the all group. For example:
    a. Consider the comparisons between complementary subgroups. That is, it may be more appropriate to compare a racial subgroup to the white/Caucasian subgroup to not double count students. Similarly for ELL, SwD, and Low SES subgroups, it may be more appropriate to use the non-subgroup as a comparison (Low SES vs. non-Low SES).
    b. While subtracting proficiency rates is common, it may not be the most accurate representation of closing the gap. Alternatives could include the percentage of the gap closed (i.e., % proficient of low SES/ % proficient of non-low SES) or comparing a measure with more variability (e.g., average scale score within grade or z-score differences between groups).
    c. Super subgroups may still be a viable approach for high-stakes to avoid duplicate counts, with individual subgroups reporting of achievement gaps being used for reporting. A potential approach for justification could be to compare the predictive power of super-subgroups vs. an aggregate of multiple subgroup comparisons to show the differences between the two. Alternatively, if achievement gap data are shown to be highly collinear (e.g., multiple subgroup gaps), that may serve as justification to decrease the proportion individual achievement gaps serve in this indicator and within the accountability system.

Again, returning to the policy and technical considerations, practitioners should ensure that the way in which growth or gap calculations are conducted serve both the policy and technical goals of the system. Are the measures reflecting progress as intended by the selected indicators? Do they demonstrate sufficient variability to differentiate schools? Are they exhibiting enough equity that they are not unduly influenced by the composition of the students within a school or district?

English Language Proficiency

The inclusion of English Language Proficiency poses a different kind of problem in that it may not be a ubiquitous measure for all schools—particularly in states with a homogenous population. English Language Proficiency (ELP) rates have traditionally been reported at the district level using Title III

Annual Measureable Achievement Objectives (AMAOs). However, it appears that this requirement speaks to an increased focus in equity of educational access and performance.

States should begin by establishing a minimum n size for subgroup identification. Smaller n sizes typically allow as many students as possible to be taken into account for a school's performance. Another outcome may include identifying more schools with students who would take an ELP assessment. States should conduct exploratory analyses to determine the percentage of schools with subgroups, including those students who would take an ELP assessment, using various n sizes in alignment with their theory of action. However, several considerations should be made with regard to this indicator.

1. Because of the difference between ESSA and NCLB with regard to ELP, it may be prudent to assign this indicator a smaller proportion of the accountability system than other indicators to avoid benefitting or negatively impacting schools.
2. The overall accountability system could be treated as a ratio of possible points to be accumulated. This would allow the denominator to adjust accordingly, taking into account the contextual differences of schools in a state (i.e., fewer total possible points in the denominator for schools without students taking an ELP assessment).
3. For particularly homogenous states with few ELP students, it may be worth pushing to advocate for a LEA-level accountability measure, with disaggregated reporting for schools with sufficient students who take an ELP assessment. Alternatively, states will need to consider treating schools with and schools without ELP subgroups as using different calculation approaches.

Practitioners will need to be especially attuned to issues around equity, availability, reliability, and comparability. Given the potentially different concentrations of English Learners, it may be difficult to make comparisons across schools. Further, depending on the transiency of the state population or the exit patterns of ELs, the consistency of data will be a factor to consider. States may also want to consider identifying ways to award credit to schools that are able to exit students from English Language Learner status to avoid unintended consequences associated with deflating EL proficiency rates.

Graduation Rate

Graduation rates for accountability have historically relied on the 4 year adjusted cohort rate (ACR). However, accountability workbooks and later Requests for Flexibility from ESEA have also utilized the 5 and 6-year ACRs. These may be a more appropriate indicator of graduation given the shift of post-secondary governing bodies to use a 4 and 6-year graduation timeline for community colleges and 4-year institutions, respectively. The following approaches could be considered when determining how to use graduation rates in an accountability system.

1. Calculate 4-, 5-, and 6-year ACRs and examine the distributions of each to identify which ones reflect the state's theory of action and, if appropriate, the most variability to support differentiation.
2. Graduation rates often rely on state-specific policies that may make direct comparisons difficult, thus making multi-state recommendations problematic. Depending on the stringency of

graduation polices, states may want to establish normative goals based on a certain percentiles of performance to establish graduation rate targets.

In examining graduation rates, the selection of the ACR timeframe should be clearly aligned to the theory of action. That is, there should be a justification that states can identify to justify the selection. By examining longitudinal data associated with K12 and post-secondary performance (if available), practitioners may be able to accumulate evidence that speaks to the appropriateness of using a 5- or 6-year ACR.

Aggregating Indicators

While the selection of indicators is somewhat contextualized, considering how these indicators interact may have a more generalized set of steps. That is, using simulated data, one can approximate the impact on how each set of indicators and their relative weights differentiate schools.

However, because there is no "best number" that differentiates school quality, the relative weights of the accountability system should be selected theoretically—in this case, by using the state's theory of action. The initial weights and indicators can then be checked analytically to maximize differentiation using the stated theory of action. As a starting point, the following steps can be taken using simulated data.

1. Conduct exploratory analyses to determine how the measures correlate with each other and the outcome to ensure different facets of school quality are being represented. If indicators are collinear, they may need to be adjusted.
2. Conduct exploratory analyses (e.g., factor analysis) to understand whether and to what degree all of the indicators actually load on the outcome. If an indicator does not load on the outcome, determine whether this was done by design (i.e., expanding the construct of school quality) or if it is due to some unforeseen circumstance.
3. Conduct a regression analysis with the predictors and determine the partial variance accounted for to determine an estimated impact of the measures selected. One could then verify that the partial variances reflect the weights (i.e., indicating the relative impact of each indicator). Depending on results
    a. The weights could be adjusted to better reflect their associated predictive weights, or
    b. The indicator measure could be modified to increase its predictive value.
4. Once weights are established, extreme cases of schools should be identified to determine whether the model tolerates outliers (e.g., high growth, very low achievement, vice versa). This can then also be used to compare against the stated goals in the theory of action to determine whether there is a violation of stated principles.

The aggregating of indicators becomes an exercise in balance. A system that prioritizes differentiation of schools but does so with complicated weighting structures may be considered more accurate, but may also be more difficult when trying to garner buy-in. The approaches that are mentioned above are sample suggested analyses. The primary driver should be the theory of action, which may stipulate how

indicators are aggregated. If so, these and other analyses can help practitioners gather evidence that the accountability system is accurate and reflective of the stated goals.

Appendix Z

In Search of Unicorns: Considering the "5th Indicator" in ESSA Accountability

Scott Marion and Susan Lyons, Center for Assessment

June 30, 2016

The Elementary and Secondary Education Act (ESEA) was reauthorized as the Every Student Succeeds Act (ESSA) this past December. Most of the indicators required under ESSA are at least familiar, even if the specific metrics proposed may be new under ESSA. However, the types of metrics and indicators suggested for the "fifth indicator" are relatively new and generally have not been used in accountability systems. My colleague, Chris Domaleski, refers to this as the "unicorn indicator" because it is something we've all heard about, but rarely (never) have seen. The specific passage from the law defining this indicator follows:

(v)(I) For all public schools in the State, not less than one indicator of school quality or student 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.

As can be seen above, there are several psychometric characteristics required of this indicator—valid, reliable, and must differentiate performance—but, in general, the options for what can be used as an indicator are fairly wide open. That said, it will be important to consider each of these technical requirements as one thinks about potential indicators. While reliability is easily defined, the validity of

an indicator (within a system context) is less clear but needs to be based on a well-articulated theory of action. Current thinking about "differentiate" is that the law intends for indicators to have a fair amount of true variability among schools compared with indicators such as elementary school attendance that essentially acts as a constant in the system.

Conceptualizing the Other Measure of School Quality

States need to be thoughtful about this additional indicator regarding how it fits with their conceptions of educational accountability and school quality. Do state leaders think this additional indicator will 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 precursors to the achievement and growth academic indicators? For example, some might want to include an indicator of student engagement because they think it is a precursor to higher levels of student achievement, while certain social-emotional learning indicators help broaden our characterizations of school quality. Obviously, there is considerable overlap among these conceptions. We expand these conceptualizations below. We argue that this indicator can be characterized along the following dimensions:

- ✓ School or student unit of analysis
- ✓ Level of inference
- ✓ Potential corruptibility
- ✓ Level of data burden

School quality or student success

This is likely the most important dimension on which to conceptualize the fifth indicator. School quality indicators are intended to broaden the ways in which we characterize school effectiveness beyond the typical indicators of reading and mathematics achievement. Generally, the data are collected through individual students, but the school is the unit of analysis. On the hand, data for student success indicators are collected and reported at the student level. Table 1 below summarizes this distinction.

Indicators of School Quality	Indicators of Student Success
➢ 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 (specific change agents) ➢ Causal precursors to the student success indicators.	➢ Value propositions about what characterizes a successful student. Broadens definition of successful student. ➢ May be a precursor skill to an achievement, growth, or attainment indicator ➢ Outcomes-based ➢ Drive innovation through local flexibility ➢ Multiple viable pathways for achieving student success and those pathways may be context-specific

Table 1. Conceptualizing indicators of school quality and student success.

To help bring this contrast to light, we offer several examples of potential indicators in each of the two categories in Table 2 below.

Indicators of School Quality	Indicators of Student Success
➢ Attendance ➢ Achievement gap ➢ School climate ➢ Percentage of students entering STEM field ➢ Percentage of students in extra-curricular activities ➢ Percentage of students enrolled in an art course ➢ Percentage of all students with access to career pathways ➢ Percentage of all students who earn college credit while in HS ➢ Educator quality (qualifications, experience, effectiveness) ➢ Suspensions/expulsions ➢ Quality of local assessments or assessment practices ➢ Engagement in professional capacity building	➢ Truancy ➢ Student engagement ➢ Persistence/dropout rate ➢ Data drawn from post-secondary outcomes ➢ Social-emotional skills ➢ Physical fitness assessment results ➢ Credits earned by end of ninth grade ➢ Algebra readiness by the end of 7th grade ➢ Enrollment in advanced coursework ➢ Earning a career/technology certificate ➢ Earning college credit ➢ Persistence in post-secondary education

Table 2. Examples of indicators of school quality and student success.

These are just examples and, as can be seen, certain types of indicators can be used at the student or school level or both, depending on how one would like the data reported and used.

Level of inference

The level of inference associated with the indicator is an important dimension and one that interacts with the next dimension we discuss; potential corruptibility. A high inference indicator might be something like school climate where 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. As one can see, there are multiple steps along this inferential chain that must be validated to substantiate the claim that one is indeed measuring school climate. Student engagement might be a student success indicator that falls along the higher inference end of the continuum because even if it is operationalized as something like attendance because strong inferences are required to support the notion that attendance is an accurate indicator of true 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 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.

Potential 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 through surveys or other similar approaches. We must carefully consider "Campbell's Law" when using any indicator, but especially those that are easily corruptible if they are used as part of a high stakes (or at least publicly reported) accountability systems.

> *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*[19].

While over 40 years old, this "law" has been well documented far too often since 1976, but 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 really thoughtful and careful in our accountability system design and especially in the design of this fifth indicator.

One of the ways we may minimize the 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 3-4 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 that required clear demonstrations of evidence where corruption may be minimized such as some of the lower inference indicators discussed above.

Level of 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 low on the data burden end of the 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 lower-skilled workers raising some data quality concerns. We are not

[19] Campbell, Donald T., Assessing the Impact of Planned Social Change. The Public Affairs Center, Dartmouth College, Hanover New Hampshire, USA. December, 1976.

necessarily advocating for considering only measures with low data burden, but we argue that such burden needs to be considered carefully.

Validation and Theories of Action

The considerations and dimensions discussed above are important, because it highlights how one approaches the development and validation of the indicator. 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. For example, one might think of a school engagement initiative that encourages students to participate in community service or other applied projects. Such engagement may be thought to help students hone leadership skills and other characteristics associated with being responsible global citizens, which are not measured well on tests. It stands to reason, then, that validating the indicator with assessment data would be misplaced. Rather, we would seek other data thought to affirm our understanding of the construct. On the other hand, one might operate from a perspective 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 is less certain. Importantly, these conceptions are both potential hypotheses that should be clearly articulated in a theory of action and then evaluated empirically as data are collected.

Effective accountability systems need to be designed according to a well-articulated theory of action that clearly lays out the intended goals and outcomes as well as proximal and intermediate indicators and the mechanisms and processes necessary to realize these goals. This is a critical first step in designing the system, but it is especially critical when selecting/creating an indicator or indicators in this category. Further, Hargreaves and Braun[20] offered concrete suggestion for designing improvement-based (compared with punitive) accountability system. While the design of the entire system should attend to these recommendations, the selection of the 5th indicator (or indicators) offers opportunities to try to meet these improvement intentions. Accountability systems should incentivize the types of behaviors state and district leaders want to see and disincentivize the behaviors they do not want to see. The fifth indicator needs to be thought about in this light.

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 the accountability system sits. Therefore, states should considering using the ESSA accountability system to support behaviors aligned with larger system goals. For example, if a key goal is to ensure that all students have legitimate access to higher education, a 5th indicator could be percentage of students who have enrolled in and received credit for dual enrollment and/or Advanced Placement classes.

[20] Hargreaves, A. & Braun, H. (2013). *Data-Driven Improvement and Accountability.* Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/publication/data-driven-improvement-accountability/.

Several of the indicators provided as examples, such as student and educator engagement, may be related to the example goal mentioned above, but they may too indirect to incentivize the increased use of college coursework in high schools that such a state may hope to see on a large scale. Before considering potential indicators or types of indicators that may support such goals, it is important to remember the ESSA requirements related to this indicator. In particular, the statute requires "the same indicator or indicators [to be] used for each grade span," which limits the flexibility a state has in tailoring different indicators to different school districts depending on need. A narrow reading of the law suggests that the same measures (e.g., school climate survey, assessment of student engagement) must be used at least for each grade span. However, a more flexible reading would suggest as long as the same indicator is used (e.g., student engagement), schools/districts may be able to use different specific measures of the same indicator. While this may be addressed in the rules, states should consider making a case for this latter position if they can document some degree of comparability across schools. It appears that Senator Lamar Alexander, the primary author of ESSA, supports this more flexible approach.

As discussed above, states should consider the differences between low and high inference indicators as well as indicators that may be distal proxies to the behaviors the state hopes to see. School climate is an example of a high-inference indicator, where the characterization of the indicator is usually based on combining information from surveys and/or interviews of members from various stakeholder communities. On the other hand, the type of higher education preparatory indicator discussed above appears to be a lower inference indicator because accountability leaders need to create rules for counting preparation coursework. On the other hand, such an indicator still rests on an assumption that successfully completing dual enrollment or AP classes leads to better outcomes in college. In this case, one could simply turn to the research literature to see if such an assumption is supported. Needless to say, such literature reviews should be part of considering any potential 5th indicator.

These examples of potential fifth indicators should be tied to supporting state-articulated goals for school improvement and student learning. 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. Further, states should use the time prior to 2017-2018 to try out a variety of indicators to evaluate the quality of data received and the burden associated with collecting such data.

65x25 Work Group Recommendations

June 27, 2016

1156 15th Street NW, Suite 850, Washington, DC 20002
202.547.2222 (phone) • 202.547.2224 (fax)

www.HCMStrategists.com

Align. Advocate. Advance.

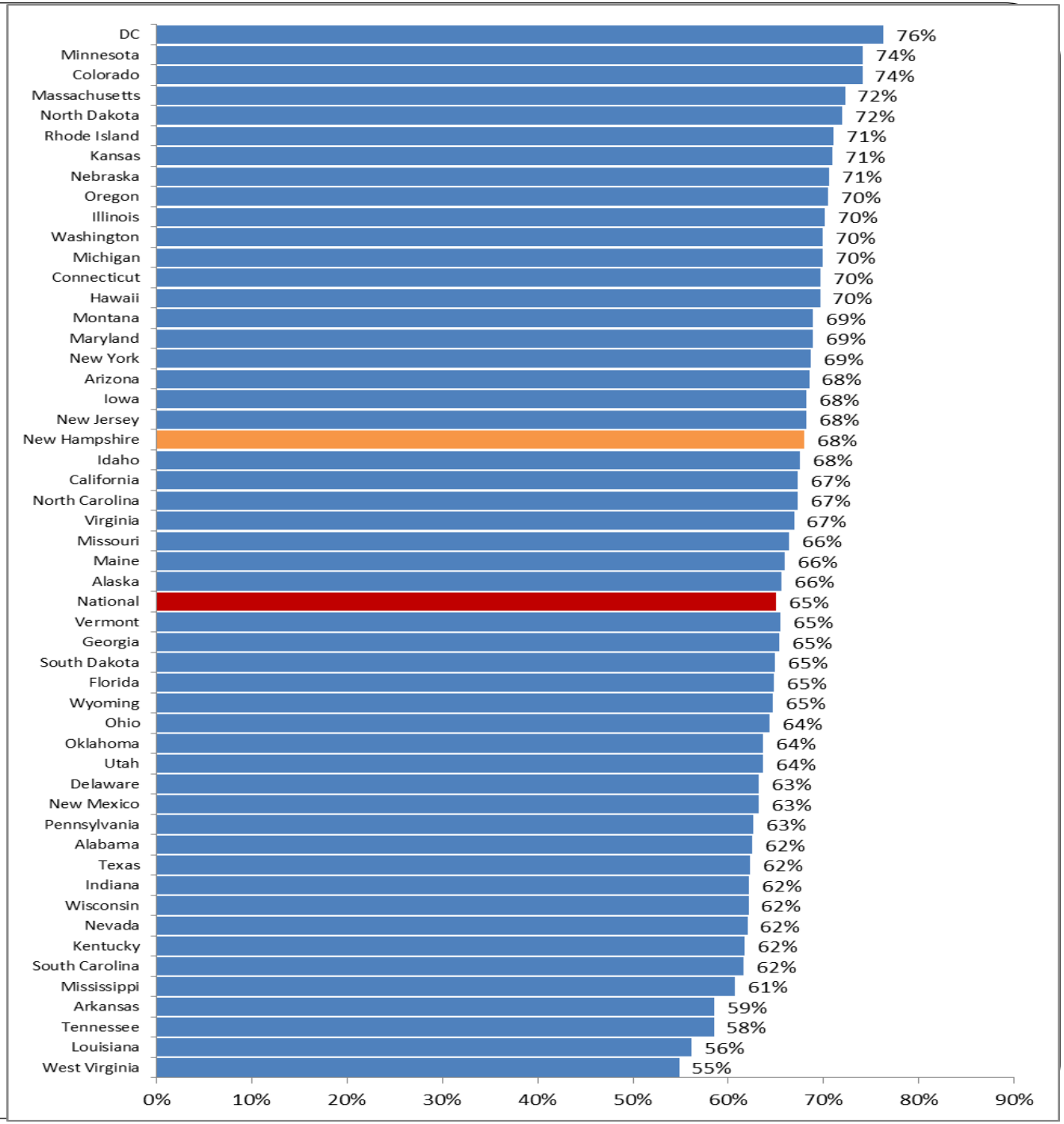
OBJECTIVE

- Share recommendations on the following for the Coalition's consideration and approval:
  - Definition of a high-quality postsecondary credential
  - Target populations and number of credentials to get to 65x25
  - Strategies to reach 65x25
  - Structure to support and sustain statewide effort to reach 65x25

The Result & Goal Statements

- New Hampshire's citizens have the education necessary to meet their life goals, as well as the current and future economic needs of the state
- 65% of 25-64 year olds will have a high-quality postsecondary credential by 2025

68% of NH jobs will require postsecondary education – compared to the national estimate of 65%



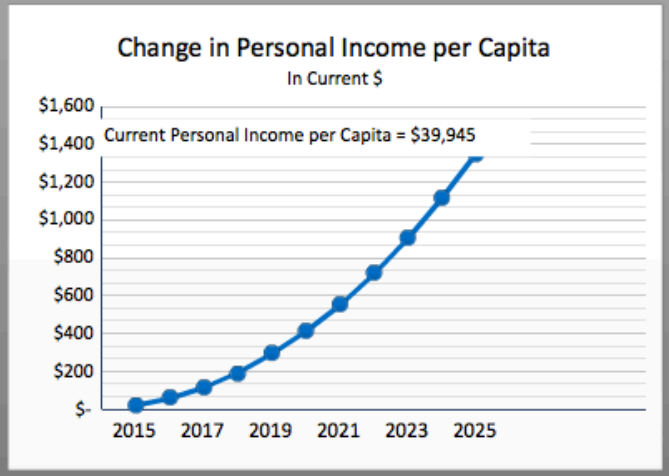
GEORGETOWN UNIVERSITY
Center on Education and the Workforce

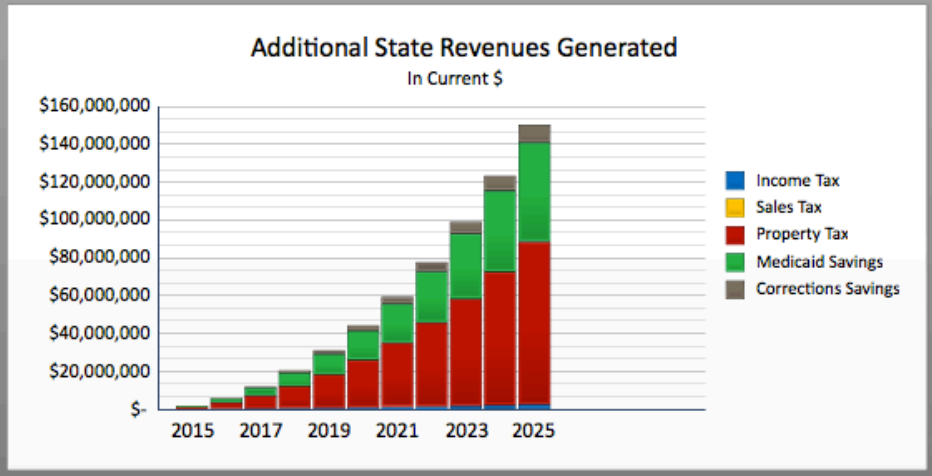
Return on Investment after Reaching Goal

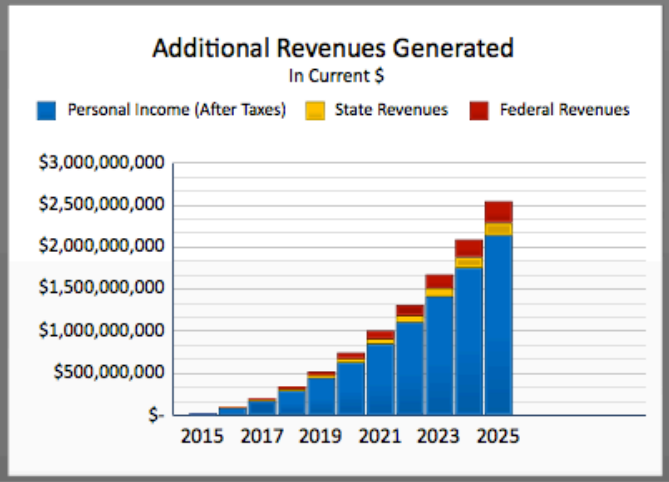
Calculating the Economic Value of Increasing College Credentials by 2025
New Hampshire

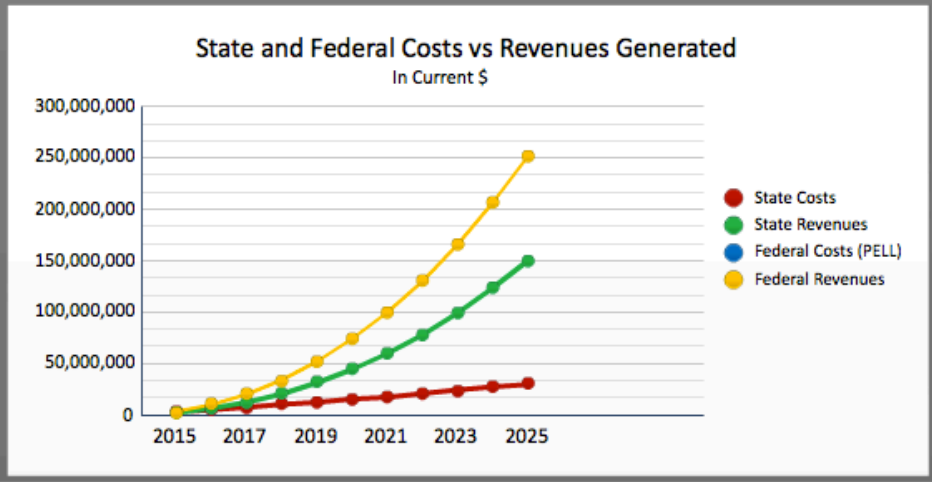
Start Over | Save Scenario | Print Page

Increase College Attainm... | View Returns on Investm... | Instructions and Definitio...





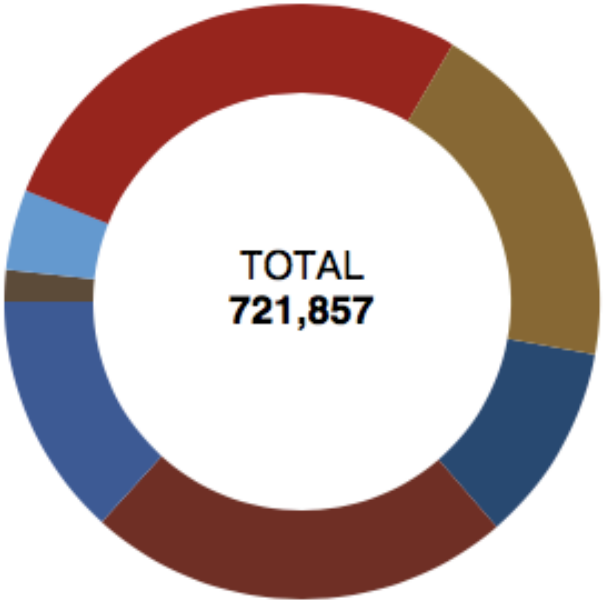




Note: The default positions reflect current rates and values. The results in 2025 assume linear progress toward goals.

Created by NCHEMS and CLASP

Levels of education for New Hampshire residents, ages 25-64



Less than ninth grade	12,390	1.72%
Ninth to 12th grade, no diploma	31,578	4.37%
High school graduate*	197,436	27.35%
Some college, no degree	140,021	19.40%
Associate degree	78,096	10.82%
Bachelor's degree	166,619	23.08%
Graduate or professional degree	95,717	13.26%

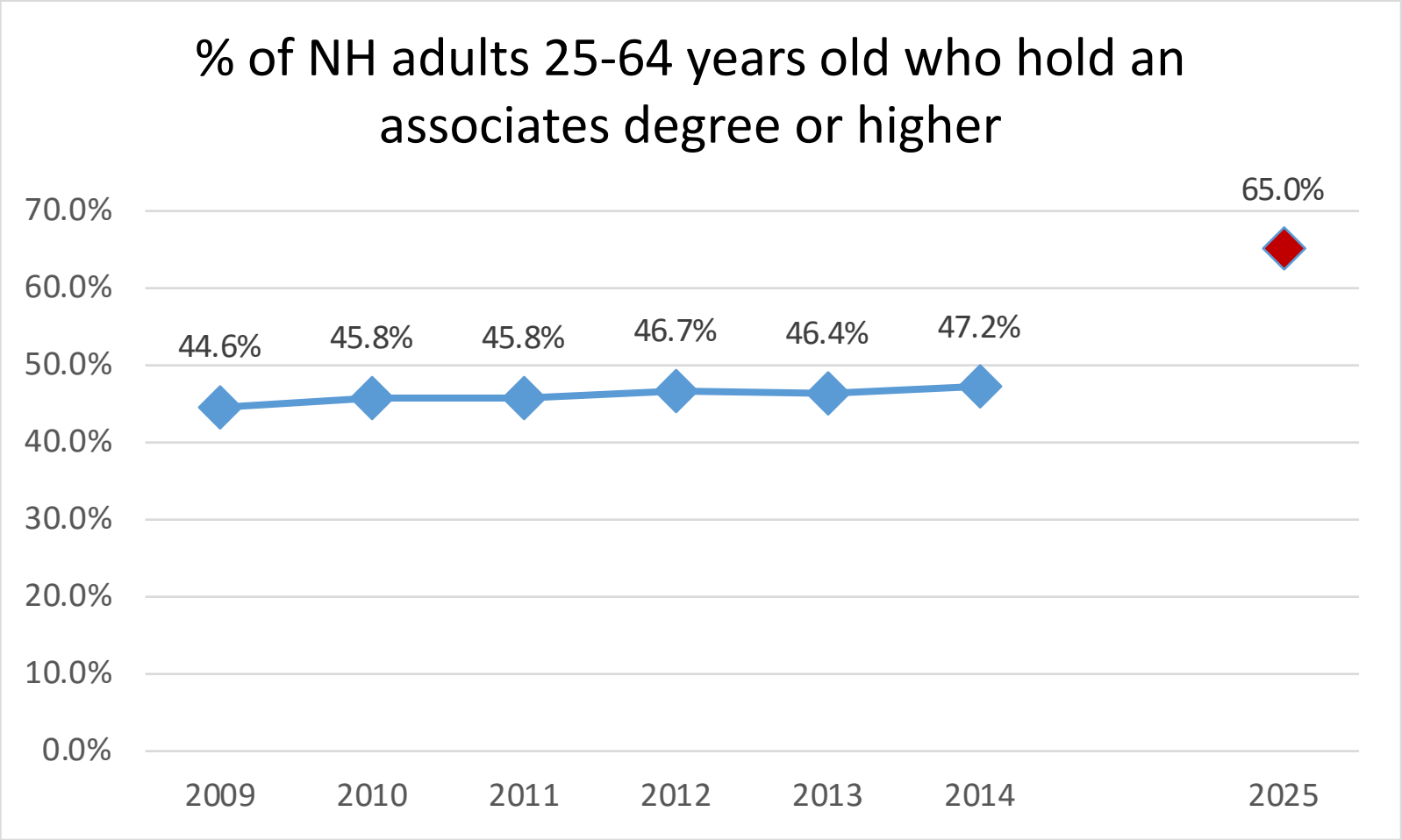
Estimated attainment of certificates: **2%**

** including equivalency*

Source: U.S. Census Bureau, 2014 American Community Survey

*Note: The accompanying pie chart does not account for residents who have earned high-value postsecondary certificates. The percentage on the right – admittedly, an **estimate** – aims to fill that gap. To calculate this percentage, labor market experts at the Georgetown University Center on Education and the Workforce used Survey of Income Program Participation 2008 Wave 12 data (2012) and data from the Integrated Postsecondary Education Data System (IPEDS) 2014.*

Progress Toward the Goal



Source: American Community Survey

Defining "High-Quality Postsecondary Credential"

- A "high quality credential" is one that inspires an individual to deeper learning in a subject of their choosing, is valued by employers, and will be beneficial in the individual's career or personal pursuits.

- *Finer Point*
  - New Hampshire Center for Public Policy Studies plans to work with the BIA to survey employers regarding what is valued.
  - Additional surveys could be conducted to better understand needs of non-profits and individuals.
  - Target timeframe for BIA survey is summer 2016.

Measures of Success

- Progress Measures (previously approved)
  - Supply (credentialed workers) : Demand (jobs requiring postsecondary education) ratio
  - Postsecondary Enrollment (% of high school graduates enrolling immediately after graduating)
  - In-state Enrollment (% of resident students remaining in-state for college)
  - Postsecondary completions (completions per 100 full-time equivalents)
  - Proportion of 25-49 year olds enrolled (as percent of 25-49 year olds without a Bachelor's degree)

- Develop next level of progress metrics
  - New Hampshire Center for Public Policy Studies intends to convene a working group to begin to identify state workforce development metrics that align with other statewide efforts.
  - K-12 Pathways workgroup to develop measures to monitor progress toward every K-12 student obtaining a career related credential or degree

Number of Credentialed Adults Needed in NH by 2025: 454,019

	2014	2025
Working-age adults (25-64 years old)	721,857	698,491
Number of adults w/ associates or higher	340,432	65%
Number of certificate holders	40,451	
Total # of adults with credentials	380,883	454,019
	53%	65%

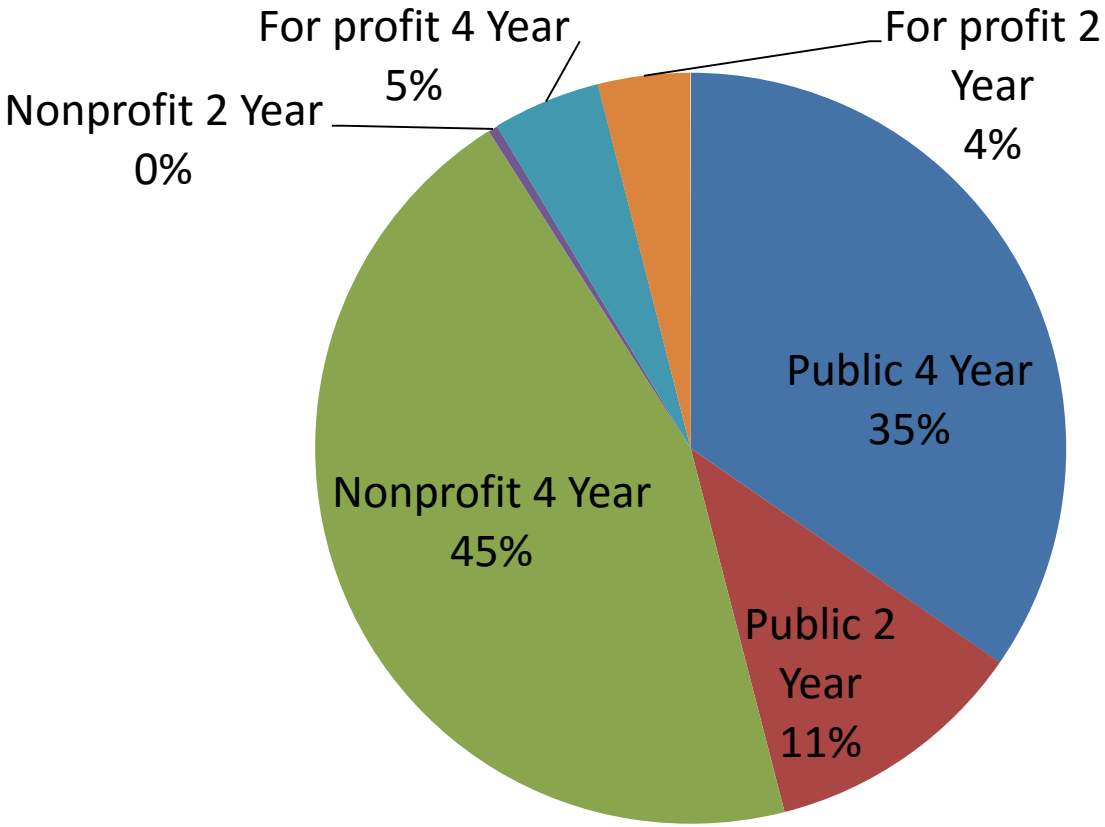
Source: Census, NHOEP, IPEDS

Number of ADDITIONAL Credentials Needed in NH by 2025: 83,819

Number of credentialed individuals needed in NH in 2025	454,019
Estimated number of credentialed residents in 2025 at current attainment rate (53% of 698,491)	370,200
Total additional credentials NH needs to produce and retain by 2025	83,819

Source: Census, NHOEP, IPEDS

2013-14 NH Credentials Awarded by Sector: 20,768



Source: IPEDS

Breaking Down 84,000 Credentials

- How does NH begin to operationalize 84,000 to make it an actionable amount?

- By Target Population
  - Who are the populations that effect 84,000?
    - **Adults**
      - Approx. 340,000 adults with HS diploma or some college, no degree
    - **Traditional-age students**
      - Steadily declining population
    - **Migration**
      - Need to attract and retain adults with high quality postsecondary credentials

Breaking Down 84,000 Credentials

- As one *potential* graphical representation, NH would need to increase the number of adults with high quality credentials by 1,867 per year to reach 84,000 by 2025
  - Assumes each year that NH minimally produces the same number as the previous year and produces an additional 1,867 credentials

									1,867	
								1,867	1,867	
							1,867	1,867	1,867	
						1,867	1,867	1,867	1,867	
					1,867	1,867	1,867	1,867	1,867	
				1,867	1,867	1,867	1,867	1,867	1,867	
			1,867	1,867	1,867	1,867	1,867	1,867	1,867	
		1,867	1,867	1,867	1,867	1,867	1,867	1,867	1,867	
	1,867	1,867	1,867	1,867	1,867	1,867	1,867	1,867	1,867	**Grand Total**
Annual Total	**1,867**	**3,733**	**5,600**	**7,467**	**9,333**	**11,200**	**13,067**	**14,933**	**16,800**	**84,000**
Year	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**2025**	

Breaking It Down by Target Population

	Total	Annual Production (2017-25)
New Credentials Needed	84,000	1,867
Target Populations:		
Adults	34,000	756
Traditional Age	31,000	689
Net In-Migration*	19,000	2,111

* This is an annual average in-migration of 25-64 year olds with a high-quality postsecondary credential (simply 19,000 divided by 9 years remaining until 2025). The methodology differs from the annual figures for adults and traditional age students, where the amount displayed is compounded each year.

Strategies to Achieve 65x25

- Strong focus on engaging and supporting adults, in particular those with some college, no credential
  - Not possible to achieve the goal with traditional age students alone

- Strategies will require effort across multiple partners:
  - High schools / CTE
  - 2- and 4-year postsecondary institutions; public and private
  - Employers
  - Government entities
  - Non-profit organizations

ACHIEVING 65x25

Goal	Strategies	Target Group	HS / CTE	2-year HE	4-year HE	Non-profits	Employers	Gov't
1. Increase number of adults with high quality credentials	• Adoption of focused prior learning assessments (PLA) • Outreach to adults that started but did not complete a degree ("non-completers") in NH • Outreach to non-completers within 2- and 4-year institutions	Adults with some college, no credential		X	X		X	
2. Attract and retain more individuals with high quality credentials in NH	• Link students to internships, practicums and apprenticeships • Support programs targeting young adults to live and work in NH • Support efforts to enroll more out-of-state students • Identify efforts to attract adults with high quality credentials to live and work in NH	HS and college students	X	X	X	X	X	X
3. Increase employer participation in efforts that support employees with continuing their education and completing their degree	• Explore, plan and implement an employer initiative that supports this goal, such as Next Step Maine or other similar effort.	Adults		X	X		X	

Achieving 65x25 (cont.)

Goal	Strategies	Target Group	HS / CTE	2-year HE	4-year HE	Non-profits	Employers	Gov't
4. Improve career pathways	• Develop sector-specific workforce education pathways	Adults with some college, no degree	X	X	X		X	X
5. Increase number of NH high school students attending college in NH	• Facilitate achievement of initial credentials in HS • Support dual enrollment programs (e.g., Running Start, early college programs) • Expand competency-based learning • Expand experiential learning	NH high school students	X	X	X		X	

6. Improve postsecondary persistence and completion	• Implement policies and programs that reduce time to credential and removes barriers to completion, including:
	  • Guided pathways
	  • Block scheduling
	  • Structured schedules
	  • Full-time is 15 credits
	  • Co-requisite remediation courses
	• Implement outreach program to contact students who have failed to register on-time
	• Develop robust transfer programs
		All students with focus on working adults/non-traditional students/low skilled workers		X	X			

ACHIEVING 65X25 (CONT.)

Goal	Strategies	Target Group	HS / CTE	2-year HE	4-year HE	Non-profits	Employers	Gov't
7. Increase affordability	• Student loan forgiveness partnerships for employees/ educational benefit • Support scholarships for target groups • Support increase in state funding for higher education	Adults with no credential; traditional age NH students		X	X		X	X

Structural Support to Sustain 65x25

- A visible and credible business-led initiative that:
  - Has a steering committee comprised of representatives from partner entities in order to set overall goals and strategies for implementing recommendations from the 65x25 work groups
  - Has staff able to carry out strategies
  - Includes a public report-out mechanism to inform on progress toward goals
  - Develops and implements a communications/marketing plan
  - Is a source for collecting and interpreting data
  - Has sufficient corporate/foundation support to hire staff and secure data collection

Recommendations of the 65x25 Workgroup

- To approve:
  - Definition of a high-quality postsecondary credential
  - Target populations and number of credentials to get to 65x25
  - Strategies to reach 65x25
  - Structure to support and sustain statewide effort to reach 65x25