A Theory of Action for the ETS K-12 Assessment Portfolio Components

E. CAROLINE WYLIE, ETS

Abstract

Most K-12 assessment systems integrate summative, interim (including both benchmark assessments and testlets), and formative assessment components. This report provides a theory of action for the ETS K-12 Assessment Portfolio. A theory of action illustrates the claims made about a program through a logic model (a diagram that links program components to intermediate and long-term outcomes for various stakeholders) and a review of supporting literature for those claims. We first briefly describe the assessment components of the system and some associated professional support, and then we identify five principles that undergird the theory of action and the logic model. We then present the logic model diagram and set of claims in the logic model that are ultimately intended to lead to improved student outcomes in the educational system. The report ends with a summary of logical, theoretical, and empirical evidence that supports these claims.

Keywords: K-12 assessment system, summative assessment, interim assessment, formative assessment, theory of action

Corresponding author: E. Caroline Wylie, E-mail: ecwylie@ets.org

Table of Contents

A Theory of Action for the ETS K-12 Assessment Portfolio Components

The design team focused on creating a coherent set of assessments that would provide information to stakeholders at all levels of the educational system and that would facilitate decisions for improving teaching and learning for all students. The top priority for the new set of assessment components was that if used together they would demonstrably improve teaching and learning in U.S. classrooms. The ETS K-12 Assessment Portfolio components could be used together in a full system or separately as part of other systems. For the purpose of the theory of action we will assume that the components are being used together. To the extent that this is true in a specific context will impact the extent to which this proposed theory of action will apply.

The ETS K-12 Assessment Portfolio components include summative, interim (including both benchmark assessments and testlets), and formative assessment components, initially focused on mathematics and English language arts (ELA) in Grades 3–8 and high school. This assessment system is intended to make what students know and can do more visible to parents/guardians, students, teachers, and administrators. The system is intended to use each state's standards in conjunction with research-based learning progressions and key practices not just to identify students' knowledge, skills, and abilities but also to provide additional information about how students got there and what the appropriate next instructional steps might be.

This report provides a preliminary theory of action for an assessment system using all of the ETS K-12 Assessment Portfolio components. A theory of action illustrates the claims made about a program through a logic model (a diagram that links program components to intermediate and long-term outcomes for various stakeholders) and a review of supporting literature for those claims (Leusner & Lyon, 2008).

Given that the ETS K-12 Assessment Portfolio components is still in development, there are no ETS-specific empirical studies from which we can draw evidence about the program's ability to improve instruction and student learning. However, we can examine research literature on summative, interim, and formative assessment to identify the plausibility of our claims where other systems or assessments have similar features. This report sets out how the ETS K-12 Assessment Portfolio components are intended to improve teaching and learning if the combination of the summative, interim, and formative assessment components together with professional supports is used in a sustained and supported manner. As the program develops and matures, ETS-specific evidence will be collected to support and/or refine the theory of action.

In the following sections, we first briefly describe the assessment and professional support components of the system, and then we identify five principles that undergird the theory of action and the logic model. We then present the logic model diagram and set of claims in the logic model that are ultimately intended to lead to improved student outcomes in the educational system. The report ends with a summary of logical, theoretical, and empirical evidence that supports these claims.

ETS K-12 Assessment Portfolio Components

The ETS K-12 Assessment Portfolio components include summative, interim (including both benchmark assessments and testlets), and formative components. They were developed using the same set of standards and a unique combination of learning progressions. We recognize that while students engage in all three assessment types, the information that comes from each one may vary in utility to different actors in the system:

1. The *summative assessment* is primarily intended to measure students' proficiency with respect to each state's standards in ELA and mathematics in the initial launch (with science anticipated for future versions) for accountability purposes. At launch, the summative assessment will be an online linear test, after which it will use a multistage adaptive approach to tailor the assessment to student ability. In addition, individual student year-to-year growth is intended to be monitored using a vertical scale, which allows for comparisons to be made across grade levels. Secondarily, the assessment will provide more specific characterizations of level of understanding for individuals and groups in selected subdomains as starting points for receiving classroom teacher and administrator follow-up. In each content area (mathematics and ELA), the online assessment will be built to balance efficiency with rigorous assessment of complex constructs by using multistage adaptive testing, using a wide range of technology-enhanced items and utilizing both human scoring and automated methods for the scoring of constructed-response items. The results (total scaled score, subscores, and proficiency levels for all students and for subgroups) will be available in an interactive reporting system that can be used by state, district, and school administrators to examine student progress toward college and career readiness in mathematics and ELA.
2. *Interim assessments* are intended to be used by teachers for instructional modification and by administrators for monitoring student learning between annual summative administrations. We separated the various roles that interim assessment can play into two distinct components: benchmark assessments and testlets. Interim *benchmark assessments* provide snapshots at the start, middle, and end of the year and are intended to be predictive of student performance on the summative assessment, recognizing that the predictive power of the benchmark assessment increases the closer in time that it is taken to the summative. They can be used primarily by teachers and administrators as monitoring tools, because each benchmark assessment is built to the same blueprint as the summative assessment, albeit shorter. The *interim testlet assessments* will be developed to target critical aspects of standards, replacing what is often referred to in schools as *common assessments*. They can be used at flexible intervals as best determined by the needs of the classroom teacher as a pretest for baseline information at the start of a unit or after a unit of instruction has been completed, such as at the start or end of a unit focused on proportional reasoning in mathematics or argumentation in ELA. They will provide results that highlight progress on critical learning progressions and key practices. The testlet results can help teachers determine additional opportunities to reinforce ideas and concepts that students may not have fully mastered. The testlet results, mapped to targeted learning progressions, also provide a way for teachers to calibrate qualitatively their own classroom assessment judgments.
3. The *formative assessment* component of the system is the part that teachers and students may interact with most frequently, and over which the teachers will have the greatest level of control, to support teachers' and students' ongoing formative assessment processes. Teachers will be able to use the formative assessment resources that model good teaching and learning practices to support ongoing instruction, assessing students' competency development along a continuum of development from novice to expert. These resources will differ from the interim assessment components in that they are intended to provide finer grained information (compared to the testlets) for teachers and students. In the final product, there may be a range of formative resources, from quick checks-for-understanding to professional learning supports for teachers to help engage with formative assessment practices that inform immediate adjustments to teaching and learning. The formative assessment component will support student-to-student interactions (e.g., peer review, collaboration, metacognition, peer tutoring) that research has demonstrated can have powerful impacts on student learning (Black & Wiliam, 1998). Information in the formative setting will focus on helping teachers make sense of student responses in light of the standards, learning progressions, and key practices and will provide

guidance to the next instructional steps. While testlets support periodic assessment, the formative assessment resources will support teachers' and students' ongoing, daily assessment of learning.

4. While not an assessment component, an important aspect of the ETS K-12 Assessment Portfolio components will be the *system supports*. System supports are intended to encompass a wide range of resources beyond the manuals and online resources that support test administrations. System supports also include resources that support assessment literacy and that help teachers understand and make use of the learning progressions and key practices both as part of assessment reports and as teaching and learning supports. In particular, these resources are intended to target hard-to-teach topics in College and Career Ready standards, such as argumentation in ELA or mathematical modeling, and to enable teachers to make appropriate use of the learning progressions and key practices in their instruction, not just as part of the assessment system. We plan to provide resources that will model good teaching and learning practices, supported by just-in-time professional development opportunities through webinars, online videos, and resources for school-based professional learning communities.
5. The intention is that each of the ETS K-12 Assessment Portfolio components are developed to be a coherent part of the rest of the system. Coherence comes in part from the use of the learning progressions and key practices that both inform development of and reporting for each component and that are part of the resources provided in the system supports for teachers to inform classroom instruction. Coherence also is a part of system development so that the student experience is similar across the assessment components, with consistent display of items and tasks and availability of accessibility and accommodation features. Results also will be available across components so that teachers can, for example, easily refer back to the previous year's summative results while reviewing testlet results. Results from one component may point to the need for further exploration of student understanding using other assessments. Whereas the total score or subscores on the summative may provide a general sense of a student's (or group's) strengths or weaknesses, the interim testlets or formative assessment components will provide a more targeted view for that student or group. The degree to which reports from one assessment component can explicitly direct teachers to other follow-up assessments or resources, and the utility of that information, will be part of the research agenda.

Principles Undergirding the Theory of Action

Five principles guide the theory of action:

1. When used together, or as part of a coherent system the ETS K-12 Assessment Portfolio components are intended to demonstrably improve teaching and learning in U.S. classrooms by supporting more informed decision making at the state and local levels, improved teacher instructional decision making, and student engagement with meaningful assessment practices. The goal is for each component of the assessment system to support valid interpretations appropriate for the intended purpose of that component, provide reliable score information for all students, and be fair in the sense of both reducing construct-irrelevant variance and making available a broad range of accommodations and accessibility features so that all students are able to demonstrate their knowledge and skill without hindrance. In addition, beyond assessing student proficiency, the system must directly support positive changes to teaching and learning in classrooms. Achieving these goals requires a constellation of approaches, starting at the beginning of the assessment system design process, grounded in the following concepts:
    - *Accessible*. Accessibility is optimized through applying universal design principles (Dolan & Hall, 2001) and by providing a robust set of accessibility features, accommodations, and

resources across the system for all students, and in particular for students who typically have access challenges, such as students with disabilities and students who are English learners.

- *Engaging*. Assessment designs that increase student engagement.
- *Deeper learning*. The breadth of the standards is measured with a range of item types and tasks in the summative assessment and other components to avoid unintended consequences of teachers narrowing the curriculum to focus only on what is assessed.
- *Informative*. Reports are clear and easily understood so that student strengths and weaknesses can be readily identified by all stakeholders, including students (at the higher grade levels), parents/guardians, teachers, school administrators, and relevant officials at the district and state levels. The reports should allow meaningful disaggregation of data by subgroups to inform school, district, and state interventions and supports.
- *Demonstrable*. Good teaching and learning practices are promoted throughout the assessment system, even in the summative assessment, through strategies that enhance student understanding of expectations (e.g., the provision of criteria for success on constructed-response items in the summative assessment), tools to support student self-assessment, and targeted resources to provide additional opportunities to learn important ideas for students.
- *Educative*. Resources for teachers are provided to support the strengthening of teaching practices (e.g., professional development resources that unpack the learning progressions and key practices or resources that support assessment literacy).

2. ETS K-12 Assessment Portfolio components will draw on learning progressions and key practices to inform assessment design, reporting, and interpretation of results. Learning progressions and key practices are theories of students' competency development that articulate how learning develops from a novice understanding or practice of a competency to a more sophisticated understanding or practice. Learning progressions describe conceptual thinking and understanding at increasing levels of sophistication, often identifying major shifts in understanding. Learning progressions articulate common ways in which student understanding matures, although they may not apply to all students and should not be taken as prescriptive. Learning progressions tend to span medium to large time periods as students grapple with complex domains of knowledge (Confrey et al., 2009; Deane et al., 2012; Heritage, 2008; Smith et al., 2004). Key practices for ELA were articulated as a way of organizing the information from multiple learning progressions into a set of closely related tasks and interactions that serve a common purpose, that exercise a common set of skills (in reading, writing, and/or critical thinking), and that must be mastered to achieve college and career readiness, such as the key practice of Building and Sharing Knowledge (Deane et al., 2015). Researchers at ETS have conducted significant research on learning progressions and key practices (for additional information, see Bennett, 2010; Graf, 2009; Graf & van Rijn, 2016; van Rijn et al., 2014). Standards provide learning goals but do not articulate conceptual increments along the path to achieving those goals, nor do they include how individual skills and/or actions are configured around instructional activities. Using learning progressions at all levels of the system supports both the development of assessments that will provide information about the transition to competency and the reporting that directs teachers toward next instructional steps, particularly for those components that are closest to classroom practice.

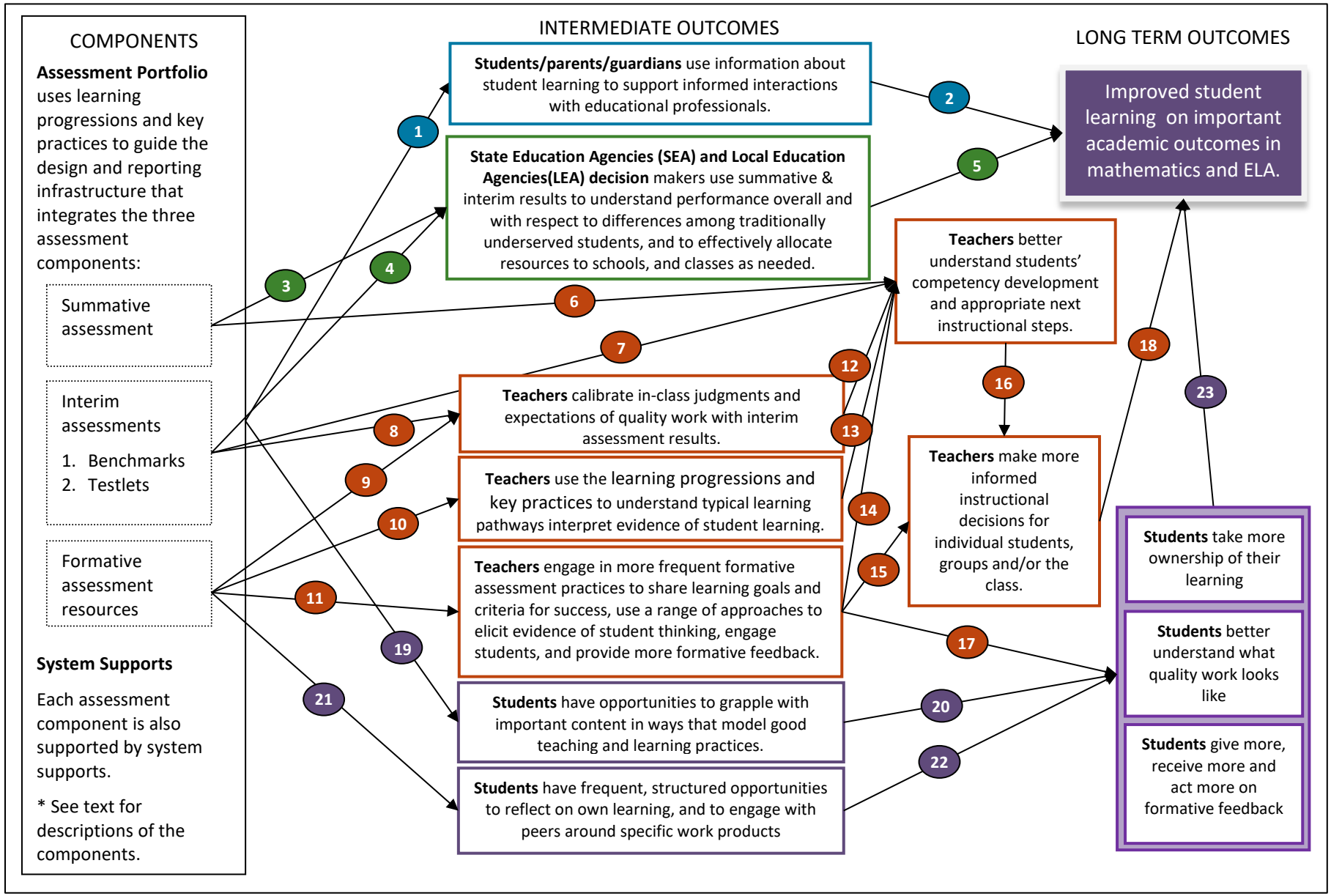
3. The reporting and feedback from the system components will address the needs of different stakeholders within the educational system who require information at different grain sizes to inform a wide range of decisions, from policy and resource decisions to instructional and learning decisions.
4. Teachers and students have complementary roles in the formative assessment process so that assessment is not just *of* students but done *with* and *by* students. The system will encourage students' engagement in self- and peer assessment to support the development of both intrapersonal and interpersonal skills in addition to cognitive skills. Teachers provide a critical role within the assessment system by interpreting assessment information in the context of their own classroom observations and modifying instruction to meet students' learning needs.
5. The ETS K-12 Assessment Portfolio components, whether used together, or as contributing parts of a coherent system will be used within a larger set of contexts, including policy, economics, educational (e.g., curriculum, instructional materials, and resources), and a learning and social contexts. The outcomes of any system (more or less positive) will depend in part on these contexts.

In the following section, we present the logic model diagram (Figure 1), in which each numbered arrow refers to a specific claim that is articulated in the text that follows the diagram. The logic model should be read from left to right. It starts on the left with the primary components of the system. To the right are the intermediate outcomes that we anticipate will occur as a result of sustained and coherent implementation of the system. These intermediate outcomes form a network of claims that are ultimately intended to lead to the long-term goal of improved student academic outcomes in mathematics and ELA. The report ends with a summary of some of the logical and preliminary empirical evidence that supports key sets of these claims.

The numbered arrows in Figure 1 represent claims about relationships between the system components and intermediate outcomes or between intermediate and longer term outcomes.

Figure 1

A logic model for the ETS K-12 Assessment Portfolio components



For Students, Parents, and Guardians

1. When students and parents/guardians have access to information about student learning from the ETS K-12 Assessment Portfolio components, they are able to have informed interactions with educational professionals.
2. When communication of information is improved between students, parents/guardians, and educational professionals, there is a positive impact on student learning outcomes.

For School, District, and State-Level Administrators

3–4. When state education agency (SEA) and local education agency (LEA) decision makers have access to valid and reliable information about student learning from the ETS K-12 Summative and Interim assessments, they are able to allocate resources more effectively at the school and classroom levels.

5. When SEA and LEA decision makers allocate resources more appropriately at the district, school, and classroom levels, student learning is enhanced.

For Teachers

6–7. When teachers have valid summative (from the previous year) and interim (from the current school year) information from the ETS K-12 Assessments about what students know in terms of key learning progressions, they better understand students' competency development and what might be appropriate next instructional steps.

8–9. When the ETS K-12 Interim assessment is used to provide periodic high-level information against both content standards and key learning progressions, teachers are better able to calibrate their in-class judgments and expectations of quality work with the interim assessment results.

10. When teachers are provided with quality formative assessment ETS resources (learning progressions and key practices, tasks, interpretive guides, professional learning opportunities) that are created to support interpretations of students' conceptual understanding, they use the learning progressions and key practices to appropriately interpret evidence of student learning, both within the assessment tasks and in other contexts.

11. When teachers are provided with quality formative assessment ETS resources, teachers engage with a range of formative assessment practices and engage their students in those practices.

12–14. When teachers calibrate their classroom judgments with the ETS interim results, use research-based learning progressions to interpret evidence of student learning, and engage in a broad range of formative assessment practices, teachers better understand student competency development to design next instructional steps.

15–16. When teachers better understand student competency development and use resources from the ETS formative component to support their formative assessment practices, they make more informed next instructional decisions for individual students, groups, and/or the class.

17. When teachers use resources from the ETS formative component to support their formative assessment practices, students are more engaged, better understand what

quality work looks like, take more ownership of their own learning, and engage more with feedback in terms of both providing feedback to peers and acting on feedback from peers and teachers.

18. When teachers tailor instruction to the specific needs of individual students, groups of students, or the whole class based on evidence, student learning improves.

For Students

19–20. When students engage with all of the ETS K-12 Assessment Portfolio components, they have opportunities to grapple with important content that models good teaching and learning practices, which in turn improves students' ownership of their learning and their understanding of expectations for quality work.

21. When ETS formative assessment resources are designed to model good instructional practices, students are encouraged to engage in reflective practices on their own learning and to provide feedback to peers using carefully designed structures.

22. When students reflect on their learning, and that of peers, students better understand what quality work looks like, take more ownership of their own learning, and engage more with feedback, in terms of both providing feedback to peers and acting on feedback from peers and teachers.

23. When students better understand expectations for quality work, take more ownership of their own learning, and engage more with feedback, student learning in mathematics and ELA improves.

Supporting Literature for the Claims Within the Theory of Action

As noted previously, the role of the logic model at this stage in the development of the ETS K-12 Assessment Portfolio components is to represent the set of claims that we are building the system to achieve. We can look to existing research studies to support the logic and plausibility of these claims and to identify gaps where we may wish to target initial research. In the following sections, we describe sets of claims from the logic model, organized first by the key users of the system and then, in some cases, broken down into smaller sets.

Claims 1 and 2: Impact on Students and Parents/Guardians

Claims 1 and 2 (blue box and arrows) in the theory of action argue that when students and parents/guardians have access to appropriate information (both student results and guides to interpreting the various forms of information) about student learning, they are able to have more informed interactions with school-based educational professionals, and that more engagement of students and parents/guardians leads to improved student learning outcomes.

One aspect of this claim focuses on the provision of information to parents/guardians, how that leads to greater engagement in their children's education, and the impact that it has on student learning. One implication within these claims is that parents/guardians value the information provided in the reports. Interestingly, several studies point toward a greater value placed on standardized testing, and resulting reports, by African American and Hispanic parents (e.g., Brewer et al., 2014; Phi Delta Kappa & Gallup, 2015; Tompson et al., 2013). The Associated Press–NORC Center for Public Affairs Research (Tompson et al., 2013) conducted a survey of parents' attitudes toward the quality of education in the United States, surveying 1,025 parents or guardians of children who completed a Grade K–12 assessment during the

2012–2013 school year. Although recent attention to the duration of standardized testing and the amount of testing may have resulted in some shifts in attitude, other findings in this report are likely to be relatively stable. Parents of color were more likely than White parents to think that it is important or extremely important to regularly assess whether students are meeting statewide expectations (85% of Hispanic parents and 82% of Black parents, compared to 69% of White parents). A similar difference is found when respondents were subdivided by income: 85% of parents earning less than $50,000 and 73% earning $50,000–$100,000, compared to 65% of parents earning more than $100,000, responded favorably to the importance of regularly assessing whether students are meeting statewide expectations. Although the survey does not attribute reasons for the different attitudes, whether by race/ethnicity or by income, it may be that specific groups of parents particularly value the importance of information provided that demonstrates whether their LEA is serving their students as well as others. Without such assessment reports, disparities cannot be identified; without identification, there can be no remediation. Although the aforementioned studies focused on the attitudes of parents toward assessment information, rather than on the actions they take as a result of the information, it is plausible to assume that informed parents/guardians are more likely to engage with educational professionals about their children's learning.

There is not a great deal of research on parents' use of score reports specifically, and what has been done has tended to focus on what parents want included in a report (A-Plus Communications, 1999), on how parents interpret information provided to them (Barber et al., 1992), or on methodologies to improve the information provided (Zapata-Rivera et al., 2012).).

One study by Munk and Bursuck (2001) that examined the perceptions of parents of high school students regarding the purposes of score reports did point specifically to the communication role of the score reports. The three most important purposes identified were (a) communicating quality of work on the curriculum, (b) communicating information about their children's achievement and work habits, and (c) communicating their children's strengths and needs with feedback on how to improve. The role of reports to support home–school communication has also been emphasized by Zapata-Rivera et al.(2013).

While not a research study, but a review of current best practices, the National Education Goals Panel (1998) explicitly recommended several approaches that could be included in score reports to support better home–school communication, such as encouraging parents to contact the teacher for more information about the test results and the inclusion of questions parents could ask of teachers to better understand the score report information and its implications for their children.

We were not able to identify specific empirical studies linking improved score reporting to improved parent–teacher engagement and from there to improved student outcomes. While studies did not target parent–teacher conversations around score reports, more generally, there is evidence that more involved parents have a positive impact on student learning. For example, two meta-analyses by Jeynes (2005, 2007) examined the influence of parental involvement on academic achievement (measured through grades, standardized tests, and other measures, such as teacher rating scales and measures of academic attitudes and behaviors) of urban elementary school children (41 studies) and urban secondary school children (52 studies). Both meta-analyses demonstrated a positive impact on student learning, with greater impact at the elementary school level. Another meta-analysis (Shute et al., 2011) of 74 studies showed some varied results, depending on which aspects of parental involvement were examined. However, there were positive relationships with student academic achievement when involvement focused on discussions between parent and child regarding academic activities and plans.

The second aspect of this claim focuses on the impact of students having access to information about their own assessment results. The original research from Dweck (1999) on growth mind-set, and more recent work in this area (Blackwell et al. , 2007; Lin-Siegler et al., 2016; Moser et al., 2011), has suggested that when students view their own learning as something they have control over, and can improve, there is a positive impact on future learning outcomes. However, no research was identified in the context of providing score reports to students, particularly middle and high school students, and so this will necessarily need to be an area of research in the future.

In summary, research has shown that parents of color and parents with lower income levels were more likely to value the information coming from summative accountability assessments. Communicating information about student results so that all parents/guardians can understand the information and take action on it is important. Research on score reports has emphasized the importance of providing information in a valid, reliable, and meaningful way (Tannenbaum, 2019), with support for parents/guardians to engage with school personnel. There is also research evidence that demonstrates that when parents/guardians are generally more involved with their children's education, there is a positive impact on student learning. It will be important to study whether and how the assessment reports are able to support parent–teacher involvement in meaningful ways, to investigate the nature of other supports and structures that may be needed for both school-level staff and parents/guardians, and to determine how assessment information can be presented to students in a way that encourages continued or improved engagement with learning.

Claims 3–5: Impact on School, District, and State-Level Administrators

Claims 3–5 (green box and arrows) in the theory of action argue that when SEA and LEA decision makers have access to reliable and valid summative and interim assessment information, they are better able to allocate supports and resources to schools and classes as needed.

The evidence for the potential positive impact that an assessment system can have on school-, district-, and state-level administrators is both logical and empirical. As part of the development, there will be a constant focus, particularly at the interim and summative levels, on identifying information for school, district, and state decision makers that will support actionable decision making. For example, particularly at the school and district levels, the use of learning progressions/key practices in reporting student learning will provide information that goes beyond proficient/not proficient, at a grain size that is smaller than "Number," "Algebra," and so forth, and that aligns more meaningfully to instruction by targeting reporting against those concepts that are most relevant to the receiving teacher at the next grade level. The learning progressions/key practices, together with a multistage adaptive testing methodology, will allow for efficient and accurate measurement of students along the student ability range of the grade level being tested. In addition, during ongoing development phases, we will explore other reporting opportunities, such as using key stroke logs to provide information about writing processes. Thus, by providing more meaningful information, we argue that decision makers will be able to make better decisions.

Decision makers at SEAs use accountability data rather than interim assessment information to inform decisions and actions at the school and district levels. Researchers have identified a tension faced by states both to work to raise student achievement for all students and to mask low performance (because the federal government can sanction schools that lack improvement; Berry & Herrington, 2012; Lee, 2010). As regulations for Every Student Success Act (Civic Impulse, 2017) are finalized, this situation may change or evolve; however, many SEAs have implemented processes to group schools according to achievement on accountability assessments and to provide resources to those schools and districts. For example, the North Carolina Department of Public Instruction identifies two types of

schools (i.e., priority and focus schools) and provides professional development for school leaders, resources to support family and community engagement, instructional coaches, and technology supports for each group (Public Schools of North Carolina, n.d.). As a second example, California has implemented a process of program improvement for schools and districts and has identified a system of supports that can be accessed for schools designated as low performing (California Department of Education, n.d.). Supports come from a regional infrastructure that can provide technical assistance. Similar decision-making processes and supports can be found for most states, although impact on student learning is more difficult to find.

There seems to be less evidence about how central office administrators use assessment information, although one study (Honig & Coburn, 2007) reported on an extensive literature review of 52 books, peer-reviewed articles, and conference papers that included empirical studies, case studies, and literature reviews (note that the study's authors identified *many* more articles but eliminated manuscripts that focused on *why* districts should or could use data to inform decisions). The study identified four features of evidence that impacted how well administrators used it: availability, accessibility, ambiguity, and credibility. This study also identified that although the concept of data-driven or evidence-driven decision making is not unfamiliar in most school districts, the use of data is influenced by factors beyond the control of assessment systems:

> Our review highlights that the actual incorporation of evidence into day-to-day district central office decisions is profoundly shaped by a host of conditions including the nature of the evidence itself, opportunities for individuals to engage in collective sense making, and the availability of professional role models that demonstrate what evidence use involves. (Honig & Coburn, 2007, p. 24)

One specific way in which LEA staff might use large-scale assessment results is to make decisions about curriculum choices or areas within a curriculum that might need additional support. Although focused specifically on the consequences of large-scale assessment and accountability for students with disabilities, a study by Ysseldyke et al.(2004) noted that the effect of the IDEA mandate that students with disabilities participate in assessment and accountability systems has been to "raise the bar." In their report, Ysseldyke et al. pointed to the use of summative assessment information to identify curriculum areas that need a targeted focus at a particular grade level to guide curriculum emphases or revisions rather than using the data to inform action for individual students.

Similarly, in support of using accountability data to guide larger grained decision making, Ingram et al.(2004) reported on data they collected from nine high schools from 1996 to 1998 to understand how data were used primarily by teachers, and to a lesser extent by school decision makers, to improve practice. They found that teachers were much more likely to identify a wider range of measures than just student achievement data as useful, including student behavior, affect, feedback, and postsecondary choices. Standardized tests were noted to provide more useful information at the school level rather than at the teacher level regarding effectiveness.

Research to date has tended to focus more on school administrators' (i.e., principals') use of data rather than considering use only by district leaders. For example, Kerr et al.(2006) conducted a series of three case studies with urban districts to identify strategies to promote data use for instructional improvements. They identified several enabling factors and summarized empirical studies to support each one: strong leadership, up-front planning for data collection and use, and strong human capacity for data-driven inquiry. One finding that echoes results from other studies was that the interim assessment data were viewed as useful by the majority of principals and district staff (81% principals found interim assessment results moderately or very useful for guiding instruction). The issue of timeliness (or lack of it) of data to inform decision making was raised in this study. As the ETS K-12

Assessment Portfolio components develop, attention will need to be paid to how quickly data can be provided to district leadership and that the data provided are informative for the kinds of decisions that are made at this level.

In summary, from a logical perspective, to the extent that the ETS K-12 Assessment Portfolio components can provide insight to SEA and LEA decision makers, we expect to see positive impacts on learning outcomes. This is an area of the system, however, that does not have as much empirical support, and it will be important to study the use of assessment data at the SEA and LEA levels.

Claims 6–18: Impact on Teachers

Claims 6–18 (orange arrows and boxes) in the theory of action lay out the intermediate impacts on teachers as a result of the use of the ETS K-12 Assessment Portfolio components. These impacts include providing access to relevant student information and supporting the use of formative assessment practices, which in turn lead to better understanding of student learning and development, combined with better instructional planning, which all leads to improved student learning outcomes. In this section, we group sets of claims or call out individual claims and provide supporting logical or empirical justifications.

Claims 6 and 7 in the theory of action argue that when teachers have better summative/interim information (in conjunction with additional information from formative assessment; see later claims), they plan instruction better. Though much has been written on the differences between formative and summative assessments (e.g., Wiliam & Black, 1996) and on ways that formative and summative assessment information can together inform instruction planning (Bennett, 2011), we are aware of no empirical studies that demonstrate the impact of teachers using summative assessment information as a starting point for instructional decision making. However, from a logical perspective, it seems reasonable to surmise that in the absence of other information, a receiving teacher at the start of a new school year would look to the results of the previous year's summative assessment to get a sense of the range of student achievement in the class in order to refine instructional plans.

More work has been done in the context of interim assessment and the impact of instructional practice. For example, Goertz et al.(2009) concluded that

> while teachers accessed and analyzed interim assessment data, we found that this information did not substantially change their instructional and assessment practice. Teachers used interim assessment results largely to decide what content to re-teach and to whom, but not to make fundamental changes in the way that this content or these students were taught. (p. 6)

More recently, a study was conducted across nine school districts and 18 case study schools (Means et al., 2009). One conclusion the study authors presented in their report was that "data from student data systems are being used in school improvement efforts but are having little effect on teachers' daily instructional decisions as evidenced in case study districts" (p. viii). They identified challenges with respect to what data are available to teachers, the timeliness of the information, organizational structures that are in place to support the use of the data, and tools to help teachers act on the data. ETS will need to attend to these challenges to meet Claims 6 and 7.

Claim 8 and 9 in the theory of action argue that teachers will be able to qualitatively calibrate their classroom judgments and expectations of quality work using the interim assessment results and that this calibration process will help teachers understand student competency development better and to make more informed, appropriate instructional decisions. A similar approach was previously recommended by Morrison et al.(1995) in the context of using standardized Assessment Units in the United Kingdom as a mechanism to support teachers in calibrating their classroom judgments:

> Many teachers charged with measuring pupil achievement against the levels of the National Curriculum feel ill-equipped to do so. Some primary school teachers, for example, lack confidence in their ability to make such decisions with accuracy; they express concern that their professional judgement of what constitutes competence at a given level in a given attainment target may not accord with that of colleagues in post-primary schools. Probationer teachers and secondary teachers whose timetable includes a variety of subjects, are charged with the daunting task of reliably assigning their pupils to levels. . . . Professional impressions of "levelness" in mathematics in Northern Ireland are disseminated through a series of 30 minute pen and paper tests (Assessment Units) which assess all of those aspects of the mathematical programme of study which lend themselves to pen and paper assessment. . . . These tests are used to communicate the standards of capable, experienced teachers (chosen from primary and post-primary schools) to all Northern Ireland teachers of mathematics. It is instructive to illustrate this "education" process through an example. Consider a primary school probationer who teaches a class of 20 pupils. Based upon homework, class tests, questioning in class and projects, this teacher feels that 16 of the pupils have mastered the level 4 material in Number. Suppose that this teacher administers the level 4 Number test to the class. Should, for example, 13 of the 16 pupils whom the teacher estimated to be level 4, exceed the level 4 cutscore on the test, the teacher could conclude that his or her judgement of level 4 achievement in Number accords with that of experienced colleagues. The teacher's professional judgment would "stand" at level 4 in Number, i.e. all 16 pupils would be assigned to level 4. (pp. 179–180)

Teacher reflection is a highly valued practice, as seen in evaluation frameworks such as the Framework for Teaching (Danielson, 2013) and in the National Board for Professional Teaching Standards assessment (National Board for Professional Teaching Standards, 2016). From research, there is evidence that when teachers have an opportunity to reflect both individually on their practice and in collaboration with colleagues in a structured way, they are able to develop new understandings about their practice and to make changes to their instructional approaches that are sustained over time (D. L. Butler et al. , 2004).

In several countries, a process of social moderation is used to help teachers calibrate classroom judgments to use those data for accountability purposes (Wyatt-Smith et al., 2010). In social moderation, teachers often review student work examples and work together to score them to develop a common understanding of the expectations embodied in standards. The process of reviewing and scoring student work provides teachers with meaningful data on which to reflect.

In the context of the ETS K-12 Assessment Portfolio components, data to support both individual and joint teacher reflection will come from the interim testlet assessments. Research will be needed to understand whether and how these data from standardized assessments, teacher discussion, and reflection support calibration of teacher understanding of the standards and the expectations to which they hold students. In summary, there is empirical evidence of the value of teacher reflection, and from a logical perspective, using results from the interim testlets to spur such reflection regarding their classroom expectations is an important aspect of ensuring coherence across the system. It will be critical both to provide supports for teachers to engage in this practice and to observe their impact.

Claims 10, 13, 16, and 18 form an argument for the role of learning progressions within the formative assessment system to support teacher interpretation about student learning, to allow teachers to gain insight into how student understanding develops, and to help teachers to use that insight to make informed instructional decisions that ultimately impact student learning in a positive way.

The ETS Testlet component in particular is drawing heavily on the body of research from the *CBAL*® learning and assessment tool research initiative, which has been ongoing at ETS for almost a decade.

This work has emphasized the disciplined development of learning progressions and key practices through reviews of empirical literature, cognitive lab studies, and expert review (Graf & van Rijn, 2016), together with empirical validation of progressions (van Rijn et al., 2014). The learning progressions and key practices have informed assessment design for both summative and formative tasks (Bennett, 2010; Deane et al., 2015; Graf & van Rijn, 2016).

While there is not yet a significant body of research on the role of learning progressions to inform teacher judgments and decision making, there is evidence that teachers who have deeper content knowledge for teaching (CKT; e.g., pedagogical content knowledge) are better able to facilitate learning based on their students' understanding and needs. Darling-Hammond and Ball (1998) noted that "knowledge of children, their ideas, and their ways of thinking is crucial to teaching for understanding" (p. 16). Teachers need *knowledge of students and content* (Hill et al., 2004) to recognize the ways in which students think—their errors, justifications, misconceptions, developmental sequences, and so forth. They also need knowledge of instructional strategies and, specifically, knowledge that allows them to identify next steps in supporting students' learning (Heritage et al., 2009). Both types of knowledge are part of the more broadly defined CKT (Ball et al., 2008). Early studies to define this domain (Ball et al., 2008; Baumert et al., 2010; Hill et al., 2005) showed a significant and positive relationship between teachers' CKT and the learning gains that the students of these teachers make. Askew and Wiliam (1995) found that "learning is more effective when common misconceptions are addressed, exposed, and discussed" (p. 8). Sztajn et al.(2012) interpreted the concept of mathematical knowledge for teaching through the lens of learning progressions (or trajectories, as they call them), defining "knowledge of content and students as knowledge of the various levels of the trajectories through which learners progress from less to more sophisticated ways of thinking" (p. 149).

While using neither the terms CKT nor learning progressions, cognitively guided instruction (CGI) drew on the underlying concepts. One early study conducted by Fennema et al. (1996) that focused on the impact of a CGI professional development program demonstrated that the participating teachers consistently ascertained what students knew by eliciting and analyzing evidence of understanding and using that information to decide on future instruction. Fennema et al. concluded that these changes were supported by the success of the professional development program for deepening teachers' understanding of the research-based model of children's thinking. Another CGI study (Carpenter et al., 2000) demonstrated that when elementary teachers understood why students struggled and had resources to develop that understanding, they were more likely to adjust instruction in ways that met students' needs. Teachers took on a new orientation toward assessing and understanding students' thinking, which helped them facilitate increases in student achievement. A number of studies using treatment and control groups have demonstrated that when teachers use evidence of student learning to adapt instruction, student learning improves (Bergan et al., 1991; Fuchs et al., 1991).

In summary, there is evidence to suggest that the kinds of insights into student learning that are captured by learning progressions can help teachers make sense of student learning. Furthermore, when teachers have access to this kind of information, they are better able to adapt their instruction to meet students' needs, which ultimately has a positive impact on student learning.

Claims 11 and 14–18 form an argument for combined use of a range of formative assessment practices that together provide teachers with insight into student understanding during instruction, which in turn informs teachers' instructional decisions, which ultimately impact student learning in a positive way.

Formative assessment practices (Council of Chief State School Officers, 2008) include the use of clear learning goals and/or criteria for success, the elicitation of evidence of student understanding, the use of student self- and peer assessment (discussed in the following section), and the provision of formative feedback. Unless teachers are explicit about the purpose of learning, students often do not understand

the purpose of a specific lesson or how it fits within a larger sequence of learning (White & Frederiksen, 1998). Both quantitative and qualitative empirical research have found that students need to understand what they are learning and how they will be assessed to support one another effectively and develop a sense of autonomy. A range of studies have demonstrated the impact of providing students with success criteria that characterize quality work, and sharing learning helped students have a positive impact on learning (Tell et al., 2000; White & Frederiksen, 1998). Research has also demonstrated the importance of teachers developing and/or selecting questions that provide quality evidence regarding student thinking and misconceptions (National Council of Teachers of Mathematics [NCTM], 2000; Wylie & Ciofalo, 2006, 2009), systematically collecting evidence from all students in the classroom (NCTM, 2000; Tobin, 1987), and deepening classroom discussions in a manner that engages more students (Marshall & William, 2006; NCTM, 2000; Tobin, 1987). When teachers engage students in these practices in a sustained way, teachers are able to collect the evidence they need to allow them to better plan instruction based upon students' current understanding (Carpenter et al.1989; Graham et al., 2012; Mevarech, 1983). As noted, more informed and targeted planning of instruction leads to improved learning outcomes. In addition, empirical research studies have demonstrated the impact that carefully developed feedback can have on student learning. Students need feedback that helps them understand both what the goals are and where their work is in relation to those goals; they need to know or learn how to close the gap in order to move learning forward, and they need time and opportunity to act on that feedback (Bangert-Drowns et al., 1991; Black & Wiliam, 1998; Carpenter et al., 1989; Patthey-Chavez et al., 2004; Ramaprasad, 1983; Sadler, 1989; Shute, 2008).

In summary, to the extent that the ETS K-12 Assessment Portfolio components can provide insight to teachers to help them use learning progressions to interpret evidence of student learning, to incorporate more frequent and sustained use of formative assessment practices, and to make informed instructional decisions, we expect to see positive impacts on learning outcomes. The Kerr et al. (2006) study described in the previous section, on how district decision makers used evidence, raised a caution for teacher use of data. Kerr et al. noted the issue of the lack of flexibility that teachers perceived to change instruction, and "given the perceived pressure to stay on pace, many teachers opted to follow the curriculum instead of the data" (p. 513). School and district assessment and instructional contexts are out of the control of the ETS K-12 Assessment Portfolio components, but it is important to be aware of the issue and to sensitively design supports (particularly for principals and district administrators) to address this issue.

Claims 19–23: Impact on Students

Claims 19–23 (purple boxes and arrows) in the theory of action focus on the student role in formative assessment. Together these claims form an argument that providing opportunities for students to reflect on their work and to both give and receive feedback from their peers will result in a range of positive intermediary effects that in turn will ultimately impact student learning in a positive way.

As part of the previous set of claims focused on the teacher's role in the use of assessment information, the role of feedback from teachers was highlighted. In addition, students can receive feedback from peers. For feedback to be effective, studies have shown that it needs to identify gaps between the desired learning goal and the student's present status, that it needs to provide actionable suggestions for how to close the gap, and that students need opportunities to address the provided suggestions (Black & Wiliam, 1998; R. Butler, 1988; Ramaprasad, 1983; Sadler, 1989; Shute, 2008). In research studies focused on the impact of peer assessment and collaborative learning, students who were provided with explicit structures for providing feedback to each other, and routines for working collaboratively, were more successful and produced superior work products (King, 1992; Mercer et al., 2004; Mercer et al., 1999; White & Frederiksen, 1998).

Research has also demonstrated in a range of contexts that frequent tests can improve student learning (Hinze et al., 2013; Paul, 2015; Rohrer & Pashler, 2010). Much of the work was done in higher education contexts (A. C. Butler, 2010; Roediger & Karpicke, 2006) but more recently has been moving toward K–12 (McDaniel et al., 2013). For example, in a series of studies, McDaniel et al. demonstrated that spaced quizzing with feedback improved student performance on subsequent assessments.

Research has also shown that students benefit from structures and opportunities to regulate their learning by engaging in self-assessment of their progress toward learning goals, reflecting on that progress, and thinking metacognitively about the process by which learning is occurring. A range of empirical studies have focused on a number of specific noncognitive outcomes that are a result of students engaging in self-assessment (which collectively results in students taking greater ownership of their learning). These outcomes have included students developing internal attributions (Cohen et al., 2003), a feeling of empowerment (McDonald & Boud, 2003), improved self-efficacy (Andrade et al., 2009), and a sense of autonomy (Brookhart et al., 2004). In addition to these noncognitive outcomes, studies have shown that when students engage in self-assessment and reflection, learning improves (Cohen et al., 2003; White & Frederiksen, 1998).

In summary, to the extent that the ETS formative assessment component and supporting resources provided to teachers can make learning goals explicit, provide tasks that are structured in ways that facilitate opportunities for students to both give and receive feedback, and encourage more student reflection on their own learning, we expect to see positive impacts on learning outcomes.

Unintended Effects and Plans to Mitigate Them

The ETS theory of action describes system components together with the anticipated intermediate and long-term outcomes that are expected from the consistent and appropriate use of those components. The system supports are intended to provide resources for the range of users of the ETS K-12 Assessment Portfolio components to deepen and develop assessment literacy skills and support appropriate use of each component. However, given that it is also possible for there to be unintended negative effects, it is important for the assessment program to consider what they might be during the design stages and to take action to mitigate them.

One possible unintended consequence is contributing to achievement gaps between students in low- and high-resourced schools because each of the assessment components are delivered online requiring students to have relatively frequent access to computers, particularly for the formative assessment resources. If the use of the ETS K-12 Assessment Portfolio components exhibits the claimed benefits, and if it is used more widely in districts with greater resources relative to districts with fewer resources, it will widen achievement gaps. While this is a potential concern, the technology gap has been closing rapidly in schools and districts. The ETS formative components that will be the most technology demanding will not be fully available until after the summative component is operational, which will allow for schools and districts to make additional progress toward closing the technology gap. According to a recent White House report (White House, 2016), “as of June 2015, the percentage of school districts with high-speed broadband in their classrooms has increased from 30 percent to 77 percent—benefitting over 20 million students and cutting the connectivity divide in half” (p. 1). Furthermore, the ConnectED Initiative is anticipated “to reach its goal of connecting 99 percent of America’s students to next-generation broadband and high-speed wireless in their schools and libraries by 2018.” It will be important for ETS to closely track progress in reliable access to high-speed wireless and adequately functional digital equipment.

Another possible unintended consequence would be if principals, other administrators, and teachers unduly focused curriculum and instruction on what will be in the summative assessment. For example, ELA classroom instruction focusing only on the types of reading on passages that have previously appeared in the assessment would result in a narrowing of the curriculum. To that end, we are broadening, as much as possible (within technological and economic constraints), the construct being assessed in the summative assessment so that we develop an assessment "worth preparing for." In addition, the other components in the assessment system, in particular, the testlets and the formative assessment tasks, will be linked through the domain and evidence models to the summative assessment but will provide opportunities for students and teachers to engage in a broader set of learning and assessment activities that are coherent with and supportive of the learning assessed in the summative assessment.

Pressure as a result of teacher or school evaluation systems (specifically the use of student assessment data as a component) has the potential for teachers to try to artificially maximize the percentage of students classified as "proficient" by focusing on the so-called bubble kids, that is, those students who are close to the proficiency cut-score, and pay less attention to those students who they believe will not clear that proficiency hurdle (e.g., Booher-Jennings, 2005). Clearly such practices are neither fair nor equitable. Our intention is that the use of a multistage adaptive model for the summative assessment will provide better measurement of students who are at the floor or ceiling of the distribution and provide greater opportunities for teachers to demonstrate their impact on student learning. Recent changes in policies around teacher and school evaluation (Sawchuk, 2016), which no longer require states to include student test scores as part of evaluation systems, may reduce the pressures that lead to these behaviors. Furthermore, as part of the system support resources, we intend to provide resources that support teachers and schools engaged in balanced assessment practices.

Owing to the high-stakes nature of the summative component, as an accountability measure, there is always the possibility that users of the system will attempt to cheat in some way, such as systematic changing of student responses or coaching while students answer. Beyond administering the assessment in third party proctored conditions, an online, multistage adaptive assessment will effectively eliminate opportunities for school or district administrators to change student responses because they will be unable to access them.

Another unintended effect would be the misuse of test scores, such as using results from only one assessment to determine student placement. Given that ETS K-12 Assessment Portfolio components serve a variety of purposes, misuse is less likely than with a single summative assessment, because the interim and formative tasks can also be included to provide a broader understanding of student knowledge and skills. However, as part of the broader set of system supports, we will provide resources to support assessment literacy for relevant stakeholders. In addition, we will provide clear explanations of the kinds of interpretations that can and cannot be made on the basis of test scores.

A final unintended consequence is that the use of a fully online assessment system that includes formative assessment tasks could result in a shift in classroom instruction to only technology-based instruction, even when it is not appropriate or best practice. To support a thoughtful and appropriate use of online resources, we will provide supplementary materials and messaging to empower teachers to use the system resources as teaching tools in a variety of contexts in the classroom.

We have identified a preliminary set of unintended consequences. As the system matures, we anticipate monitoring these issues and our approaches to addressing them as well as identifying others that need their own sets of responses. We also recognize that in addition to unintended consequences of the implementation of the assessment system as described above, there are also unintended consequences based on the context within which the assessment system resides. At the time of writing we are

experiencing the impact of COVID-19 as a specific example of an unplanned "event" has negatively impacted teaching, learning, and assessment in previously unimagined ways as K-12 schools have had to go fully online or, at best, blended. As we move forward we will need to monitor how this event has impacted technology access and use, opportunity to learn, learning environments and other areas of the K-12 system that will also impact the assessment system.

References

Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *Journal of Educational Research, 102*, 287–302.

A-Plus Communications. (1999). *Reporting results: What the public wants to know. A companion report to Education Week's Quality Counts '99.* Retrieved from http://www.edcounts.org/archive/sreports/qc99/opinion/edweekresults.pdf

Askew, M., & Wiliam, D. (1995). *Recent research in mathematics education 5–16*. London, England: HM Stationery Office.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it so special? *Journal of Teacher Education, 59*, 389–407.

Bangert-Drowns, R. L., Kulik, J. A., Kulik, C. L., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213–238.

Barber, B. L., Paris, S. G., Evans, M., & Gadsden, V. L. (1992). Policies for reporting test results to parents. *Educational Measurement: Issues and Practice, 11*, 15–20.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & Tsai, Y. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*, 133–180.

Bennett, R. E. (2010). Cognitively Based Assessment of, for, and as Learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement, 8*(2–3), 70–91.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25.

Bergan, J. R., Sladeczek, I. E., Schwartz, R. D., & Smith, A. N. (1991). Effects of a measurement and planning system on kindergartners' cognitive development and educational programming. *American Educational Research Journal, 28*, 683–714.

Berry, K. S., & Herrington, C. D. (2012). Tensions across federalism, localism, and professional autonomy: Social media and stakeholder response to increased accountability. *Educational Policy, 27*, 390–409.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246–263.

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*, 231–268.

Brewer, C., Knoeppel, R. C., & Lindle, J. C. (2014). Consequential validity of accountability policy public understanding of assessments. *Educational Policy, 29*, 711–745.

Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics, 57*, 213–227.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1118–1133.

Butler, D. L., Novak, H., Jarvis-Selinger, S., & Beckingham, B. (2004). Collaboration and self-regulation in teachers' professional development. *Teaching and Teacher Education, 20*, 435–455.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*(1), 1–14.

California Department of Education. (n.d.). *Statewide system of school support (S4)*. Retrieved from http://www.cde.ca.gov/sp/sw/ss/index.asp

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. (2000). *Cognitively guided instruction: A research-based teacher professional development program for elementary school mathematics* (Report No. 003). Madison, WI: National Center for Improving Student Learning and Achievement in Mathematics and Science, University of Wisconsin.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*, 499–531.

Civic Impulse. (2017). S. 1177 — 114th Congress: Every Student Succeeds Act. Retrieved from https://www.govtrack.us/congress/bills/114/s1177

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*, 119–142.

Confrey, J., Maloney, A., Nguyen, K., Mojica, G., & Myers, M. (2009). *Equipartitioning/splitting as a foundation of rational number reasoning using learning trajectories.* Paper presented at the 33rd conference of the International Group for the Psychology of Mathematics Education, Thessaloniki, Greece.

Council of Chief State School Officers. (2008). *Attributes of effective formative assessment: A work product coordinated by Sarah McManus, North Carolina Department of Public Instruction for the CCSSO Formative Assessment for Students and Teachers (FAST) SCASS*. Washington, DC: Author.

Danielson, C. (2013). *The Framework for Teaching: Evaluation instrument*. Princeton, NJ: Danielson Group.

Darling-Hammond, L., & Ball, D. L. (1998). *Teaching for high standards: What policymakers need to know and be able to do*. Philadelphia, PA: National Commission on Teaching and America's Future, Consortium for Policy Research in Education.

Deane, P., Sabatini, J., Feng, G., Sparks, J., Song, Y., Fowles, M., & Foley, C. (2015). *Key practices in the English language arts (ELA): Linking learning theory, assessment, and instruction* (Research Report No. RR-15-17). Princeton, NJ: Educational Testing Service.

Deane, P., Sabatini, J., & O'Reilly, T. (2012). English language arts literacy framework. Princeton, NJ: Educational Testing Service.

Dolan, R. P., & Hall, T. E. (2001). Universal design for learning: Implications for large-scale assessment. *IDA Perspectives, 27*(4), 22–25.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*, 403–434.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*, 617–641.

Goertz, M. E., Olah, L. N., & Riggan, M. (2009). *Can interim assessments be used for instructional change?* (Policy Brief No. RB-51). Philadelphia, PA: Consortium for Policy Research in Education.

Graf, E. A. (2009). *Defining mathematics competency in the service of cognitively based assessment for Grades 6 through 8* (Research Report No. RR-09-42). Princeton, NJ: Educational Testing Service.

Graf, E. A., & van Rijn, P. (2016). Learning progressions as a guide for design: Recommendations based on observations from a mathematics assessment. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of test development* (2nd ed., pp. 165–189). New York, NY: Routledge.

Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Unpublished manuscript.

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment.* Washington, DC: Council of Chief State School Officers. Retrieved from http://www.ccsso.org/Resources/Publications/Learning_Progressions_Supporting_Instruction_and_Formative_Assessment.html

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment. *Educational Measurement: Issues and Practice, 28*(3), 24–31.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371–406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal, 105*(1), 11–30.

Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language, 69*, 151–164.

Honig, M. I., & Coburn, C. (2007). Evidence-based decision making in school district central offices: Toward a policy and research agenda. *Educational Policy, 22*, 578–608.

Ingram, D., Seashore Louis, K., & Schroeder, R. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106*, 1258–1287.

Jeynes, W. H. (2005). A meta-analysis of the relation of parental involvement to urban elementary school student academic achievement. *Urban Education, 40*, 237–269.

Jeynes, W. H. (2007). The relationship between parental involvement and urban secondary school student academic achievement a meta-analysis. *Urban Education, 42*, 82–110.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112*, 496–520.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27*, 111–126.

Lee, J. (2010). Trick or treat: New ecology of education accountability system in the USA. *Journal of Education Policy, 25*(1), 73–93.

Leusner, D. M., & Lyon, C. J. (2008). *Providing a research basis for ETS products* (Research Memorandum No. RM-08-01). Princeton, NJ: Educational Testing Service.

Lin-Siegler, X., Dweck, C. S., & Cohen, G. L. (2016). Instructional interventions that motivate classroom learning. *Journal of Educational Psychology, 108*, 295–299.

Marshall, B., & William, D. (2006). *English inside the black box: Assessment for learning in the English classroom*. London: Granada Learning.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology, 27*, 360–372.

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice, 10*, 209–220.

Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). *Implementing data-informed decision making in schools: Teacher access, supports, and use*. Menlo Park, CA: SRI International.

Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal, 30*, 359–377.

Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal, 25*(1), 95–111.

Mevarech, Z. R. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics, 14*, 415–429.

Morrison, H. G., Healy, J., & Wylie, E. C. (1995). Teacher knows best: A solution to the marks-to-levels problem in National Curriculum Testing. *British Educational Research Journal, 21*, 175–182.

Moser, J. S., Schroder, H. S., Heeter, C., Moran, T. P., & Lee, Y.-H. (2011). Mind your errors: Evidence for a neural mechanism linking growth mind-set to adaptive posterror adjustments. *Psychological Science, 22*, 1484–1489.

Munk, D. D., & Bursuck, W. D. (2001). What report card grades should do and communicate. *Remedial and Special Education, 22*, 280–287.

National Board for Professional Teaching Standards. (2016). *What teachers should know and be able to do.* Arlington, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Education Goals Panel. (1998). *Talking about tests: An idea book for state leaders*. Washington, DC: U.S. Government Printing Office. Retrieved from http://govinfo.library.unt.edu/negp/reports/98talking.PDF

Patthey-Chavez, G. G., Matsumura, L. C., & Valdes, R. (2004). Investigating the process approach to writing instruction in urban middle schools. *Journal of Adolescent & Adult Literacy, 47*, 462–476.

Paul, A. M. (2015). Researchers find that frequent tests can boost learning. *Scientific American, 313*(2), 1–7.

Phi Delta Kappa & Gallup. (2015). The 47th PDK/Gallup Poll of the public's attitudes toward the public schools: Testing doesn't measure up for Americans. *Phi Delta Kappan, 97*(1). Retrieved from http://www.fsba.org/wp-content/uploads/2014/01/PDK-Gallup-Poll-2015.pdf

Public Schools of North Carolina. (n.d.). *Overview: Turning around North Carolina's lowest-achieving schools (TALAS)*. Retrieved from http://dst.ncdpi.wikispaces.net/

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*, 4–13.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*, 406–412.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.

Sawchuk, S. (2016, April 6). ESSA loosens reins on teacher evaluations, qualifications. *Education Week,* 14–15.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189.

Shute, V. J., Hansen, E. G., Underwood, J. S., & Razzouk, R. (2011). A review of the relationship between parental involvement and secondary school students' academic achievement. *Education Research International, 2011*, 1–10.

Smith, C., Wiser, M., Anderson, C. W., Krajcik, J., & Coppola, B. (2004). *Implications of research on children's learning for assessment: Matter and atomic molecular theory.* Paper commissioned by the Committee on Test Design for K-12 Science Achievement. Center for Education, National Research Council.

Sztajn, P., Confrey, G., Wilson, P. H., & Edgington, C. (2012). Learning trajectory based instruction: Toward a theory of teaching. *Educational Researcher, 41*, 147–156.

Tannenbaum, R.J. (2019). Validity aspects of score reporting. In. D. Zapata-Rivera (Ed.), *Score reporting*: *Research and applications* (pp. 9-18). New York, NY: Routledge.

Tell, C. A., Bodone, F. M., & Addie, K. L. (2000). *A framework of teacher knowledge and skills necessary in a standards-based system: Lessons from high school and university faculty.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Tobin, K. (1987). The role of wait time in higher cognitive level learning. *Review of Educational Research, 57*(1), 69–95.

Tompson, T., Benz, J., & Agiesta, J. (2013). *Parents' attitudes on the quality of education in the United States*. Retrieved from http://www.apnorc.org/projects/Pages/parents-attitudes-on-the-quality-of-education-in-the-united-states.aspx

van Rijn, P. W., Graf, E. A., & Deane, P. (2014). Empirical recovery of argumentation learning progressions in scenario-based assessments of English language arts. *Psicología Educativa, 20*, 109–115.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.

White House. (2016). *ConnectED: Realizing the promise of digital learning* [Press release]. Retrieved from https://www.whitehouse.gov/sites/default/files/docs/fact-sheet_connected_realizing-the-promise-of-digital-learning.pdf

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*, 537–548.

Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice, 17*(1), 59–75.

Wylie, C., & Ciofalo, J. (2006). *One diagnostic item: Then what?* Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, San Francisco, CA. [[Caroline: Should this first author be C. E. Wylie? If not, I'll need to add initials to the in-text citations.]]

Wylie, E. C., & Ciofalo, J. (2009). *Diagnostic items in math and science* (Final Report No. R305K040051). Washington, DC: Institute for Education Sciences.

Ysseldyke, J., Dennison, A., & Nelson, R. (2004). *Large-scale assessment and accountability systems: Positive consequences for students with disabilities* (Synthesis Report No. 51). Minneapolis, MN: National Center on Educational Outcomes.

Zapata-Rivera, D., VanWinkle, W., & Zwick, R. (2012). Applying score design principles in the design of score reports for CBAL™ teachers (ETS Research Memorandum RM-12-20). Princeton, NJ: ETS.

Zapata-Rivera, D., Vezzu, M., & Biggers, K. (2013, April 27–May 1). Supporting teacher communication with parents and students using score reports. Paper presented at the annual meeting of the American Educational Research Association (AERA), San Francisco, CA.

Suggested citation:

Wylie, E. C. (2020). *A theory of action* for a K-12 Assessment System. Princeton, NJ: ETS.

www.ets.org