THE CHALLENGES AND OPPORTUNITIES OF BALANCED SYSTEMS OF ASSESSMENT: A POLICY BRIEF

April 30, 2019

Scott Marion, Jeri Thompson, Carla Evans, Joseph Martineau, & Nathan Dadey

Center for Assessment

National Center for the Improvement of Educational Assessment[1]
Dover, New Hampshire

CENTER FOR ASSESSMENT

Center for Assessment

INSIDE

THE CHALLENGES AND OPPORTUNITIES OF BALANCED SYSTEMS OF ASSESSMENT: A POLICY BRIEF[1,2]

INTRODUCTION

The seminal publication *Knowing What Students Know: The Science and Design of Educational Assessment* (National Research Council [NRC], 2001) crystalized the appeal for balanced systems of assessment:

> *Assessments at all levels—from classroom to state—will work together in a system that is comprehensive, coherent, and continuous. In such a system, assessments would provide a variety of evidence to support educational decision making. Assessment at all levels would be linked back to the same underlying model of student learning and would provide indications of student growth over time* (p. 9).

This call for balanced assessment systems resulted from a recognition that most state summative assessments poorly served the primary purpose of assessment: improving learning and instruction. Educators understand that large-scale summative tests are far too distal from instruction, at the wrong grain size, and administered at the wrong time of year to make a difference in their daily practice (e.g. (Penuel & Shepard, 2016). Therefore, the interest in balancing systems of assessment—actually, to *re*balance these systems—was motivated by the desire to enhance the utility of assessments for improving learning and instruction as well as for monitoring, accountability, and evaluation.

Although it has been almost 20 years since the publication of *Knowing What Students Know*, there are few examples of well-functioning assessment systems. That said, we have learned many important things about designing and implementing high-quality assessment systems in the ensuing years. In this policy brief, we first review key conceptual issues regarding assessment system design and implementation. We then examine likely reasons for the paucity of balanced assessment systems in practice. We conclude by outlining an agenda to improve our understanding for designing and implementing balanced systems of assessment to enhance equitable learning and life opportunities for all students.

[1] This is an abridged version of the paper *A Tricky Balance: The Challenges and Opportunities of Balanced Systems of Assessments* (https://www.nciea.org/node/493), presented at the 2019 meeting of the National Council on Measurement in Education.

[2] We are grateful to Ted Coladarci and Chris Domaleski for their very helpful comments and suggestions. Any errors, however, are our own.

BALANCED ASSESSMENT SYSTEMS: CRITERIA AND COMPONENTS

What makes an assessment system balanced? An assessment system is balanced when the assessments in the system are *coherently* linked through clearly specified learning targets, they *comprehensively* provide multiple sources of evidence to support educational decision-making, and they *continuously* document student progress over time (NRC, 2001). These criteria—coherence, continuity, and comprehensiveness—create a powerful image of a high-quality system of assessments, rooted in a common model of learning. We also find that *utility* and *efficiency* are helpful considerations in thinking about the functioning of such systems when working with district and state leaders (Chattergoon, 2016; Chattergoon & Marion, 2016).

An assessment system is balanced when the assessments in the system are *coherently* linked through clearly specified learning targets, they *comprehensively* provide multiple sources of evidence to support educational decision-making, and they *continuously* document student progress over time (NRC, 2001).

We do not name specific assessment types (e.g., summative) or levels (e.g., district) that must be included in a system. It is not that we are waffling; rather, system components cannot be named in the abstract. System designers must rely on a well-specified theory of action to ensure that the various components of an assessment system meet the needs of the multiple users consistent with the intended uses. The theory of action should be created in a way to allow designers to examine the assessment system criteria delineated above.

Given the prominence of assessment types in discussions of balanced assessment systems, however, we offer additional thoughts on formative, interim, and summative assessments. **Formative** assessment must be inseparable from instruction and can be thought of as a bridge between instruction and classroom assessment (Heritage, 2010, Shepard, in press). The rest of the classroom assessment system—including unit-based performance tasks, extended projects, more-traditional tests, and so on—should be coherent with the formative assessment processes and must focus on shared learning targets.

Interim assessments are defined as

> *assessments administered during instruction to evaluate students' knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school, or district level. The specific interim assessment designs are driven by the purpose and intended uses, but the results of any interim assessment must be aggregable for reporting across students, occasions, or concepts.* (Perie, Marion, & Gong, 2009, p. 6)

Many believe that interim assessments should be part of a balanced assessment system, a notion likely fueled by advertising and marketing promises rather than evidence of utility. In fact, commercial interim assessments may distract educators from rich assessment opportunities and students from rich learning opportunities, thereby threatening system coherence. Thus, interim assessments are not required components of balanced assessment systems, but such assessments may play a productive role in balanced systems of assessment only if there is sufficient evidence of coherence and utility.

Summative assessments are designed to support various types of determinations (e.g., proficiency) given at the end of a defined instructional period such as a school year to evaluate students' performance against a set of learning targets for that period. The state summative assessment—because of its prominent role in accountability and reporting—typically plays a disproportionate role in most assessment systems and, further, is responsible for much of the imbalance in systems we see today. Therefore, state leaders may need to think about rebalancing the outsize role of the state test if they intend to support balanced assessment systems in their state[3].

BARRIERS TO ASSESSMENT SYSTEM DESIGN AND IMPLEMENTATION

We have examined much of the relevant literature over the past 20 years, and we see little attention to the reasons why one finds so few balanced assessment systems in practice. There are more potential barriers than we reasonably can consider here, but in light of the research literature and our experience, we believe these four interrelated influences pose key challenges to balanced assessment systems:

We believe these four interrelated influences pose key challenges to balanced assessment systems:

- The influence of politics, policy, and political boundaries on decisions pertaining to assessments;
- the influence of commercialization and proliferation of assessments;
- the lack of attention to curriculum and learning in the design of assessment systems; and
- the lack of assessment literacy at multiple levels of the system.

- The influence of politics, policy, and political boundaries on decisions pertaining to assessments;
- the influence of commercialization and proliferation of assessments;
- the lack of attention to curriculum and learning in the design of assessment systems; and
- the lack of assessment literacy at multiple levels of the system.

Politics and Policy

The challenges of assessment system design across political and ownership boundaries remain largely unaddressed. Different (and disconnected) political entities control various levels of the educational system and corresponding assessments. This is particularly true in the U.S. and likely in other decentralized contexts.

District control

A major issue with developing a balanced assessment system is determining who is in control. Most states cede some degree of control of curriculum and assessment to local school districts. States control the statewide end-of-year assessment, but little else. Similarly, district and school leaders control districtwide assessments and finer-grained schoolwide assessments. Finally, and perhaps most importantly, teachers are responsible for most classroom assessments in service of the instructional needs of their students. Assessment practices at one level of the system can compound

[3] To be clear, "summative" does not pertain to state-level tests solely; most district and classroom assessment systems include a summative component (e.g., for awarding grades or making competency determinations).

quality issues at other levels. Implementing balanced assessment systems cannot be a state-driven enterprise alone, and these political and ownership boundaries cannot be ignored.

We agree with Shepard et al. (2018) that districts are better positioned than states to be the controlling entity for balanced assessment systems. Districts typically control curriculum and instruction. Assessments in a balanced system must be designed to reflect and embody the corresponding learning goals and trajectories. Additionally, districts generally control hiring, professional development, supervision, evaluation, and many other structural components of the learning, instructional, and assessment systems. This puts districts in a much better position than states to create coherent and balanced assessment and learning systems.

States have a role: Tight and loose coupling

The original criteria outlined in *Knowing What Students Know* (NRC, 2001) for balanced assessment systems suggest a "tightly coupled system," where information flows among the various assessments in the system—from the statehouse to the classroom—to support multiple uses and users as efficiently as possible. This type of information flow is a high bar, likely beyond the capacity of most educational systems. We suggest "loosely coupled systems" may help bring about more coherence than what we see in typical state systems. A loosely-coupled system is where the state procures and directs the summative assessment, but it also purchases interim assessments tied to major aspects of the content standards (e.g., mathematical operations with fractions) that districts can use to supplement the information they get from the statewide summative assessment. Such systems have multiple levels of assessments all tied to the same learning targets and vision of learning, but the exchange of information is partially compartmentalized since the system does not get down to the level of the specific enacted curriculum. One benefit of loosely coupled systems is they help connect the state and some district assessments to the same learning targets by being designed together and created by the same assessment company.

We suggest "loosely coupled systems" may help bring about more coherence than what we see in typical state systems. A loosely-coupled system is where the state procures and directs the summative assessment, but it also purchases interim assessments tied to major aspects of the content standards (e.g., mathematical operations with fractions) that districts can use to supplement the information they get from the statewide summative assessment.

Turnover Among Policymakers and Shifting Priorities

Most state education chiefs have been in office for fewer than three years, similar to the average tenure of large-district superintendents. This turnover rate can bring frequent shifts in assessment policy priorities. Dealing with political differences is a formidable challenge, to be sure. We therefore advocate creating long-term structures, such as assessment policy documents (perhaps even legislation) based on credible public processes and/or long-serving and apolitical assessment advisory committees to mitigate the destabilizing effects of politics on assessment coherence.

Accountability

State accountability requirements can have perverse effects on the design and implementation of balanced assessment systems (e.g., Elmore, 2004; Hargreaves & Braun, 2013). In the world of assessment system design and implementation, strong accountability pressures can distract leaders from long-term strategies, such as building teachers' formative assessment skills, and instead cause educational leaders to grasp at short-term approaches, such as test preparation and products that promise a quick fix. Therefore, state leaders' first responsibility in promoting balanced assessment systems should be to critically examine existing and future policies for potential unintended consequences and work to eliminate or minimize such risks.

Therefore, state leaders' first responsibility in promoting balanced assessment systems should be to critically examine existing and future policies for potential unintended consequences and work to eliminate or minimize such risks.

The Commercialization and Proliferation of Assessments

Individuals operating at different levels of a system often purchase or develop new assessments to meet real or perceived needs without fully considering how existing assessments might meet the targeted needs and considering how new assessments can threaten the balance of the system.

The proliferation of assessments began in earnest during the No Child Left Behind years, with policies fixated on ever-increasing accountability targets. School and district leaders felt an overwhelming pressure to raise test scores, often against staggering odds. Many assessment vendors tried to help district leaders meet their goals or used misleading marketing claims appropriating the academic literature supporting formative assessment (Shepard, 2005; Martineau, 2004). Either way, there was a massive increase in interim assessments during the NCLB era that continues today (NRC, 2010; Perie et al., 2009). Not all of these interim assessments are low quality and ineffective. But because they rarely align with the enacted curriculum or other programs of improvement, interim assessments can distract educators from a deeper learning agenda (Konstantopoulos et al., 2016; Li et al., 2014). Consequently, these interim assessments also tend to operate in isolation outside of any local assessment system.

Curriculum and Balanced Assessment Systems

The role of curriculum in the design and implementation of balanced assessment system is a principal challenge emerging from the issues regarding political control discussed above. The through-line for coherence is a common vision of learning rooted in an enacted curriculum, describing how students are expected to progress from fragile to deeper levels of understanding and domain competence. The absence of a common vision of learning across districts serves as a significant barrier to state-led, and even district-led, balanced assessment systems. The lack of attention to curriculum (and learning progressions) similarly impedes the design and implementation of balanced assessment systems at both the state and district levels.

Classroom and formative assessment researchers (e.g., Shepard, 2000) were among the first to emphasize the central role of curriculum in balanced assessment systems. As Pellegrino (2006) cautioned, unless we reorient our assessment systems to focus on supporting teaching and learning, we likely will be unable to support our schools in developing "adaptive expertise" necessary for students to succeed in the 21st century. Assessment systems cannot support

The absence of a common vision of learning across districts serves as a significant barrier to state-led, and even district-led, balanced assessment systems. The lack of attention to curriculum (and learning progressions) similarly impedes the design and implementation of balanced assessment systems at both the state and district levels.

these teaching and learning processes unless each assessment is linked closely to how students are expected to learn the content and skills.

Assessment Literacy for Balanced Assessment Systems

Much of the blame for assessment system incoherence arguably falls on state, district, and school leaders—the decision-makers regarding assessment choices. The implementation of balanced assessment systems requires that both educators and leaders understand high-quality balanced assessment systems, and at all levels: classroom, district, and state. Inadequate assessment literacy among educators, administrators, and policymakers poses a significant barrier to the design and implementation of balanced assessment systems. Because districts are the appropriate locus of control for balanced assessment systems (Marion, 2018), developing the assessment literacy of its educators and leaders is critical to the design and implementation of high-quality balanced systems. Similarly, given the importance of the state assessment in balanced systems of assessment, we must attend to, and support increases in, the assessment literacy of state policy leaders as well.

MOVING TOWARD AN AGENDA FOR RESEARCH AND PRACTICE

The challenges associated with designing and implementing high-quality balanced systems of assessment make this work seem formidable, and, indeed, the field has a long way to go before high-quality balanced systems of assessment are commonplace. At least four concurrent strands of work are needed to ensure progress: conceptual, practical, research and evaluation, and policy.

1. **Conceptual.** *Knowing What Students Know* (NRC, 2001) and others (e.g., NRC, 2006, 2014) laid out high-level conceptual underpinnings of balanced assessment systems. Yet, the criteria proposed in *Knowing What Students Know* are not specific enough to inform policy and practice. For example, coherence is a key aspiration, but how coherent is coherent enough to ensure the assessment system will be balanced? Obvious incoherence is easy to uncover, but there is little guidance for evaluating and judging degrees of coherence. We need additional work on balanced assessment systems to make the criteria and other conceptual aspects more actionable and useful.

2. **Research and Evaluation.** We have great hopes for the initiatives we propose. Absent a corresponding research and evaluation structure, many of the efforts may well be one-offs. Therefore, research-practice partnerships are necessary for documenting proposed interventions so that others may learn from the work. For example, we asserted above that loosely coupled systems will improve the coherence and utility of the interim and summative components of the system. This is just one example. Such assertions must be supported by evidence, with plausible rival hypotheses and potential unintended negative consequences given due consideration. Similar efforts should accompany any of the major initiatives described above.

3. **Practical**. Researchers must partner with districts and states to find opportunities for designing and redesigning systems of assessment to serve as examples for others and to help refine the conceptual work. Also, tools and other supports for practitioners need further development. Finally, we must improve the quality, depth, and breadth of assessment literacy for multiple categories of stakeholders—a tremendous undertaking, to be sure.

Researchers must partner with districts and states to find opportunities for designing and redesigning systems of assessment to serve as examples for others and to help refine the conceptual work.

4. **Policy.** We have outlined the implementation challenges associated with balanced assessment systems and, in turn, the beginnings of a research and practice agenda for advancing the field. Without attending to the policy context in the design and implementation of assessments, observing high-quality assessment systems in practice will continue to be like searching for unicorns. This is particularly true for systems that feature a state component. Both accountability and assessment policies can constrain the implementation of balanced assessment systems. Many of the barriers we discussed above can turn into levers if addressed. For example, stabilizing the state assessment system and adjusting its footprint can allow district-level assessment systems to flourish. Additionally, designing accountability policies that do not narrowly focus on standardization and comparability may better support innovations in district assessment system systems to create more balance.

Without attending to the policy context in the design and implementation of assessments, observing high-quality assessment systems in practice will continue to be like searching for unicorns.

A CALL TO ACTION

We return to where we started. We sense an urgent need to improve the quality and usefulness of assessments. Balanced assessment systems have been proposed for meeting many needs, but we do not see enough examples of such systems in practice to serve as models for others to emulate. We identified several key challenges that explain why such assessment systems are rare, and we suggested approaches for ameliorating some of these challenges. We concluded by proposing a research and practice agenda for the Center for Assessment, our colleagues, and partners to guide this crucial work, which should allow us to look back after the next 20 years and see more progress than we have seen in the two decades since the publication of *Knowing What Students Know.*

REFERENCES

Chattergoon, R. & Marion, S.F. (2016). Not as easy as it sounds: Designing a balanced assessment system. *The State Education Standard, 16, 1*, 6-9

Elmore, R. F. (2004). Moving forward: Refining accountability systems. In Fuhrman, S. H. & Elmore, R. F. *Redesigning accountability systems for education* (pp.276-296). New York, NY: Teachers College Press

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/.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom.* Thousand Oaks, CA: SAGE Publishing.

Konstantopoulos, S., Miller, S.R., van der Ploeg, A., & Li, W. (2016) Effects of interim assessments on student achievement: Evidence from a large-scale experiment. *Journal of Research on Educational Effectiveness, 9:sup1, 188-208,* DOI: 10.1080/19345747.2015.1116031

Li, Y., Marion, S.F., Perie, M. & Gong, B. (2010). An approach for evaluating the technical quality of interim assessments. *Peabody Journal of Education, 85, 2,* 163-185

Marion, S. (2018). The opportunities and challenges of a systems approach to assessment. *Educational Measurement: Issues and Practice, 37, 1*, 45-48

Martineau, J. A. (2004). *The Effect of Construct Shift on the Results of Growth and Accountability Models.* (Doctoral Dissertation), Michigan State University, East Lansing, MI.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academies Press.

National Research Council. (2006). *Systems for state science assessment.* Washington, DC: National Academies Press.

National Research Council. (2010). *State assessment systems: Exploring best practices and innovations: Summary of two workshops.* Alexandra Beatty, Rapporteur; Committee on Best Practices for State Assessment Systems. National Research Council. Board on Testing and Assessment. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Penuel, W. R., & Shepard, L.A. (2016). *Handbook of Research on Teaching, Fifth Edition.* Gitomer, D. & Bell, C. (eds.) Washington, DC: American Educational Research Association.

Perie, M., Marion, S.F., & Gong, B. (2009). Moving towards a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28, 3,* 5-13.

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

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29, 7,* 4-14.

Shepard, L. A. (in press). Assessment for classroom teaching and learning. Workshop on Educational Assessment as Useful and Useable Evidence. National Academy of Education and American Academy of Political and Social Science. September 13-14, 2018

Shepard, L. A. (2005). Will Commercialization Enable or Destroy Formative Assessment? Paper presented at the ETS Invitational Conference, October 10-11, 2005, New York City

Shepard, L. A., Penuel, W. R., & Pellegrino, J. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice, 37, 1*, 21-34.

Center for Assessment

National Center for the Improvement of Educational Assessment[1]
Dover, New Hampshire

www.nciea.org