

Supporting Teacher Assessment Literacy: A Proposed Sequence of Learning

E. Caroline Wylie and Christine J. Lyon
Educational Testing Service

Submitted to Teachers College Record as a commentary paper
Published November 2017

Abstract

Our commentary focuses on a proposal for sequencing teacher professional learning opportunities to develop a well-rounded understanding of assessment practices and processes. We argue that developing teachers' assessment literacy with formative assessment is the best place to start because formative assessment happens most frequently and has the lowest stakes. To encourage generalization across assessment types, we also recommend that explicit connections be made between common key ideas across assessment types. These key ideas include purpose, evidence, questions/tasks, assessment quality, and interpretation of results. The paper will draw on current representations of balanced assessment systems and research on best practices for supporting teacher learning.

Acknowledgements

An early version of the ideas in this commentary was presented at the 2017 National Conference on Student Assessment, in Austin, Texas. We thank our collaborator Cristen McLean from the State Department in Oregon for the opportunity to develop these ideas through discussion with her.

Take care of the pennies, and the pounds will take care of themselves.

This British proverb is perhaps over-simplified financial advice. However, there is truth to the idea that learning to resist impulse buying and developing disciplined saving strategies in *low stakes* situations increases the likelihood of applying those same habits in *higher stakes* situations. Similarly, we argue that if teachers learn about key assessment literacy concepts in the lowest stakes context and with the most frequent type of assessment—formative assessment—their learning can be generalized to other assessment forms. However, to support generalization, the similarities between assessment types need to be explicit. In this commentary, we define various assessment types, identify common key ideas, and propose a learning sequence for teachers.

Balanced Assessment Systems

In a balanced assessment system, each assessment type reflects the same standards, albeit at different grain-sizes. Different stakeholders (state or district assessment directors, principals, teachers, parents, students) may rely more on different types, depending on their information needs, whether they need to know about individual student’s learning, or need more aggregated information to help monitor various levels within the system (Brookhart, 2013). The system is not functioning well if a stakeholder does not have the necessary information or is trying to use assessment information for a purpose for which it is not best suited.

There are five different assessment types commonly used in K-12 contexts (see Table 1). While most assessment types are administered primarily to provide information to stakeholders other than students, students are critical participants in the process of formative assessment.

Insert Table 1 here

The first four assessments are identified by a singular purpose, while curriculum-embedded performance assessments (CEPA) (Hofman, Goodwin & Kahl, 2015) can have a formative or summative purpose. Because the field is, once again, gaining interest in performance assessments (Marion & Shepard, 2017) we felt it was important to include them when considering how to best sequence teacher learning.

Multiple representations of balanced assessment systems have been proposed, each with strengths and weaknesses. Herman and Heritage (2007) illustrated how a comprehensive assessment system targets constructs of increasing grain-size that occur from minute-to-minute, daily, weekly, end of unit, end of quarter, or annually. Although not specific about what types of assessments, the representation illustrates how they move along a continuum with greater sampling of the standards and more aggregation at one end, while also showing how the student role decreases as the stakes and standards coverage increase. Perie, Marion, and Gong (2009) used a triangle with summative at the top, interim as a middle layer, and formative assessment as the base. A strength of this representation is its definition of interim assessment, but it does not provide information about the student role in assessment nor have an explicit space for classroom summative assessment. Brookhart (2013) proposed a four-quadrant representation, with one axis that runs from large scale to classroom focused and another axis that runs from formative to summative in purpose, thus creating an explicit place for classroom summative assessment, although each quadrant appears equally important for all stakeholders.

Each representation foregrounds and backgrounds different aspects of balanced assessment systems. To more fully capture the key aspects in the different forms of assessment, we propose a new representation (see Figure 1).

Insert

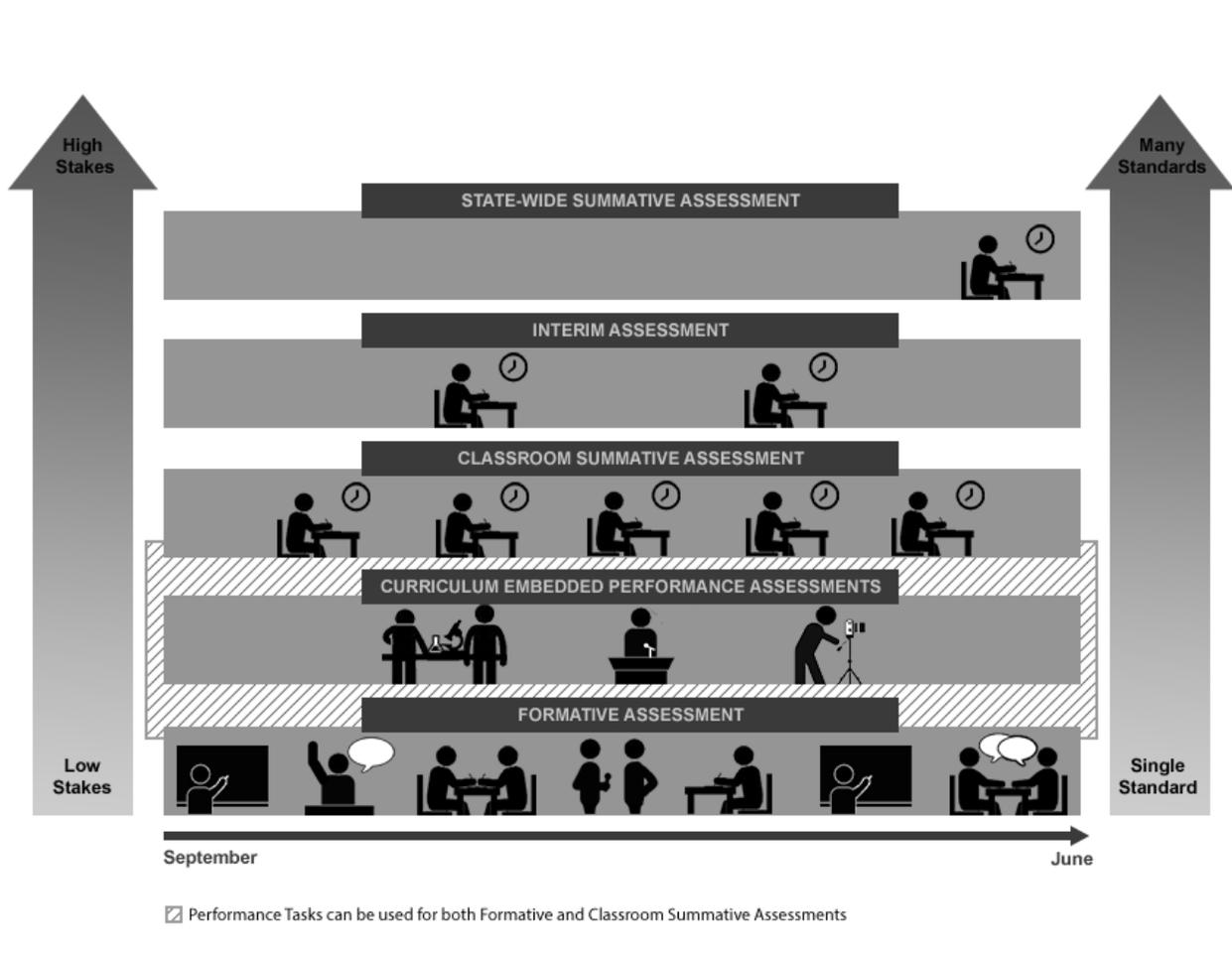


Figure 1 here

In this representation, school year is represented along the bottom and the frequency and student role in each type of assessment is illustrated pictorially. Formative assessment occurs most frequently and involves the most student collaboration (group discussions, peer feedback, public sharing of work and reasoning). Given their flexibility in use, CEPA sits between, and overlaps with both formative and classroom summative assessment, as illustrated by the shaded box. Interim assessment is used less frequently than classroom summative. Finally, statewide summative assessment takes place, once, towards the end of the school year. The arrows on the left and right of the diagram illustrate the increasing stakes and grain size of the assessment.

While this representation captures ideas across the three previous representations, it does not provide insight into how to best sequence professional learning for teachers around these

ideas. To understand this problem, we consulted the literature on effective professional learning and looked across the assessment forms to identify similarities and differences.

A Road Map for Professional Learning

Research focused on teacher professional learning has resulted in the identification of characteristics of effective professional development. For example, Garet, Porter, Desimone, Birman, and Yoon (2001) indicated that focusing on subject matter, giving teachers opportunities for “hands-on” and “collective” work, and integrating learning into the daily life of the school, are more important than focusing on other factors such as “reform” versus “traditional” professional development. Similarly, the National Staff Development Council’s report on Professional Learning (Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009) provided a review of experimental studies on professional development. While there were few well-designed studies, and causal relationships are not fully established, principles consistent with those identified by Garet et al., (2001, pp. 10-11) can be inferred. These principles include the use of professional development that is intensive, ongoing, and connected to practice; focuses on student learning and addresses the teaching of specific curriculum content; aligns with school improvement priorities and goals; and builds strong working relationships among teachers.

Regardless of the type of assessment in Figure 1, there are five common key ideas (Wiggins, 2005; Mislevy, Steinberg, & Almond, 2003):

1. Identification of what is to be measured;
2. Determining what counts as evidence of understanding;
3. Developing questions or tasks to elicit that evidence;
4. Understanding what quality evidence is; and
5. Interpreting the evidence once students have responded to the task.

Drawing on research for effective teacher professional learning and the key ideas that cut across all assessment types, we propose that the most effective and efficient way for teachers to develop assessment literacy is by first examining these five key ideas explicitly in the context of formative assessment (see Figure 2). We propose that teacher learning opportunities start with the assessment that has the lowest stakes and is closest to teaching practice (the widest, topmost layer of Figure 2), and then move to other assessment forms, spending relatively less time on each one as the frequency of use and the control that the teacher has over the design and

implementation decreases. Making the five key ideas explicit in each assessment form supports transfer of learning from one to the next.

Insert Figure 2 here

Formative assessment begins with clear learning goals and the identification of the evidence that would be necessary to demonstrate achievement of those learning goals, often called criteria for success. After identifying these goals, it is necessary to decide how to elicit evidence of understanding and then interpret that evidence in order to make decisions about next steps. This match between the learning goals and types of evidence collected goes to the heart of whether valid interpretations can be made from the evidence collected, and is similar for all assessment design although some assessments will cover more standards and require some trade-offs in terms of time and coverage.

Since formative assessment occurs most frequently, it provides opportunities for ongoing and intensive cycles of practice, reflection, and revision; it has a robust research base indicating that it has the potential to improve student learning and achievement (Black & Wiliam, 1996; Hattie & Timperly, 2007); it is developed and implemented in teachers' specific contexts; and best supported with teacher collaboration.

Finally, formative assessment also provides an opportunity to explore critical components of assessment, such as the sufficiency or trustworthiness of evidence (Stobart, 2006). The evidence that a teacher needs to guide next steps will vary based on intended use of the assessment. For example, a teacher might quickly poll students and adjust the lesson based on immediate feedback. However to plan more comprehensively for the next lesson, she might augment her observations during the lesson by asking students to identify something they understand, are puzzled by, and are curious about at the end of a lesson, or by using a longer task to collect additional evidence. The evidence must be sufficient or trustworthy for the kind of judgment being made. In this way, sufficiency of evidence is a qualitative derivative of assessment reliability in a low stakes environment where there are frequent opportunities to adjust the decisions. Additionally, in the context of formative assessment, teachers must interpret student evidence, make inferences regarding student understanding, and use those inferences to plan next instructional steps to support emerging understanding (Bennett, 2010).

Conclusion

Practicing key formative assessment ideas frequently in low stakes settings can promote teacher assessment literacy. In this commentary we argue for a move towards professional learning focused on 1) strategic, systematic development of assessment literacy, and 2) making explicit how the key features of formative assessment generalize across all assessment formats. Ideally, pre-service teacher preparation programs will examine whether and how teacher candidates develop assessment literacy, induction programs should review how assessment is addressed in the early years of a teacher's career, and in-service opportunities should be staged in a way that allow teachers to sequence their learning in meaningful ways.

References

- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74
- CCSSO (2017). Revising the Definition of Formative Assessment. Council of Chief State School Officers: Washington, DC.
- Brookhart, S. (2013). Comprehensive assessment systems in service of learning: Getting the balance right. In R. W. Lissitz (Ed.), *Informing the practice of teaching using formative and interim assessment: A systems approach* (pp. 165-184). Charlotte, North Carolina: Information Age Publishing, Inc.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American educational research journal*, 38(4), 915-945.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112
- Herman, J. L., & Heritage, M. (2007). *Moving from piecemeal to effective formative assessment practice: Moving pictures on the road to student learning*. Paper presented at the Council of Chief State School Officers Assessment Conference, Nashville, TN.
- Hofman, P., Goodwin, B., & Kahl, S. (2015). Re-Balancing Assessment: Placing Formative and Performance Assessment at the Heart of Learning and Accountability. *McREL International*.
- Marion, S. & Shepard, L. (2010). *Assessment Literacy to Support Competency-Based Education Systems and other Deeper Learning Efforts*. iNACOL Leadership Webinar June 15, 2017 28, https://www.inacol.org/wp-content/uploads/2017/06/Assessment-Literacy-for-CBE_061417.compressed.pdf
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary research and perspectives*, 1(1), 3-62.
- Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, 28(3), 5-13.

Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and Learning* (pp. 133-146). London: Sage.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. ASCD.

Table 1

Definitions of Assessment Types

<i>Type</i>	<i>Definition</i>
<i>Statewide Summative Assessments</i>	Administered by the state department of education for accountability purposes once per year. They provide evaluative data that can be aggregated and used to summarize what students have learned at the conclusion of a grade level.
<i>Interim Assessments</i>	Provide data at specific points during a school year (benchmark) or at the end of an instructional unit (common) that can be aggregated across teachers, schools, or the district to understand what students have learned.
<i>Classroom Summative Assessment</i>	A formal attempt, by an individual teacher, to determine what students have learned. The data are not generally aggregated beyond the class or teacher.
<i>Formative Assessment</i>	Ongoing classroom assessment that is an inherent part of high quality instruction. It is defined as “a planned, ongoing process used by all students and teachers during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended disciplinary learning outcomes, and support students to become more self-directed learners” (FAST SCASS, 2017).
<i>Curriculum-Embedded Performance Assessments</i>	Administered periodically to measure students’ skills based on authentic tasks that require students to demonstrate what they know, understand and can do. The evidence from curriculum-embedded performance assessments can be used to guide formative decisions during instruction or to make inferences about student understanding as a culminating activity at the end of a unit. Since they can take several class periods to complete they require a substantial investment from both teacher and students compared to other types of formative assessment and so may require specific supports for their use.

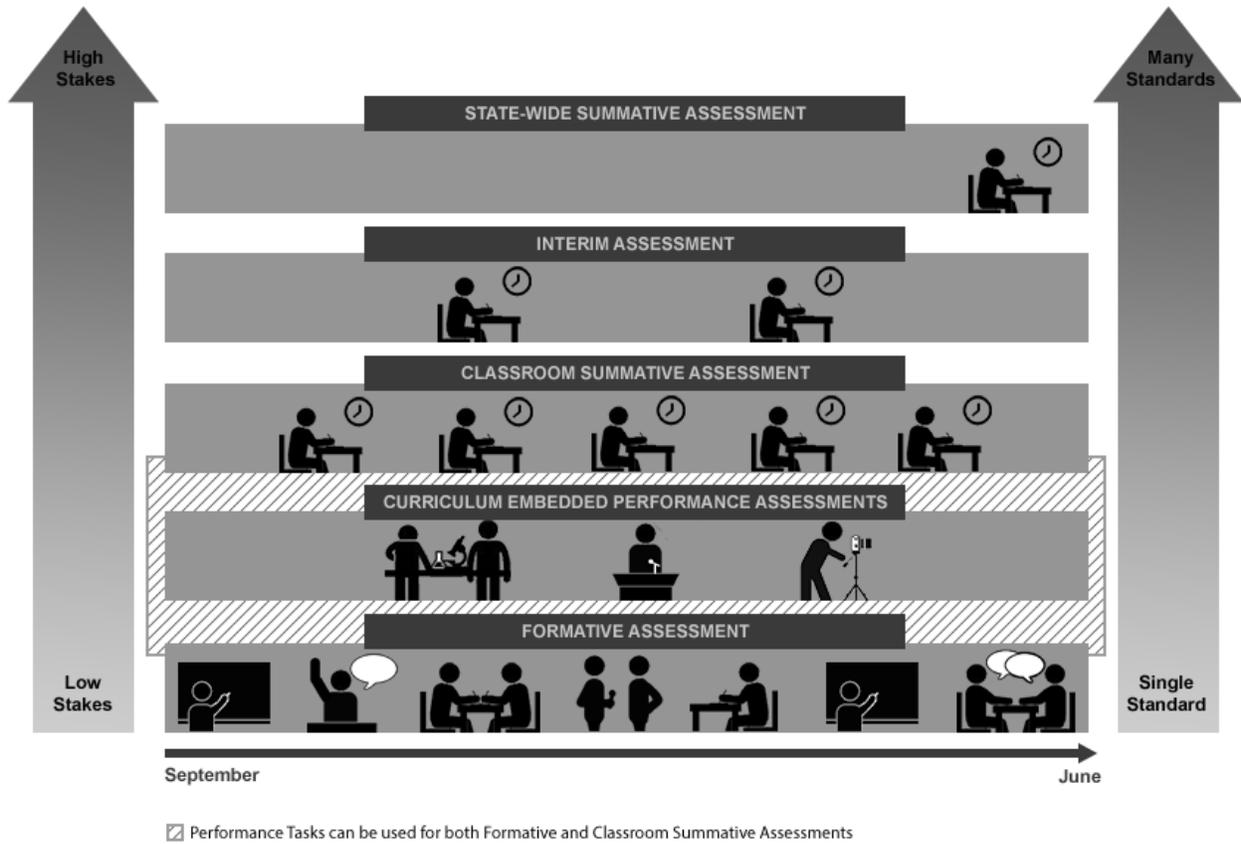


Figure 1. A balanced assessment system. Copyright © 2017 Educational Testing Service. Reprinted with permission.

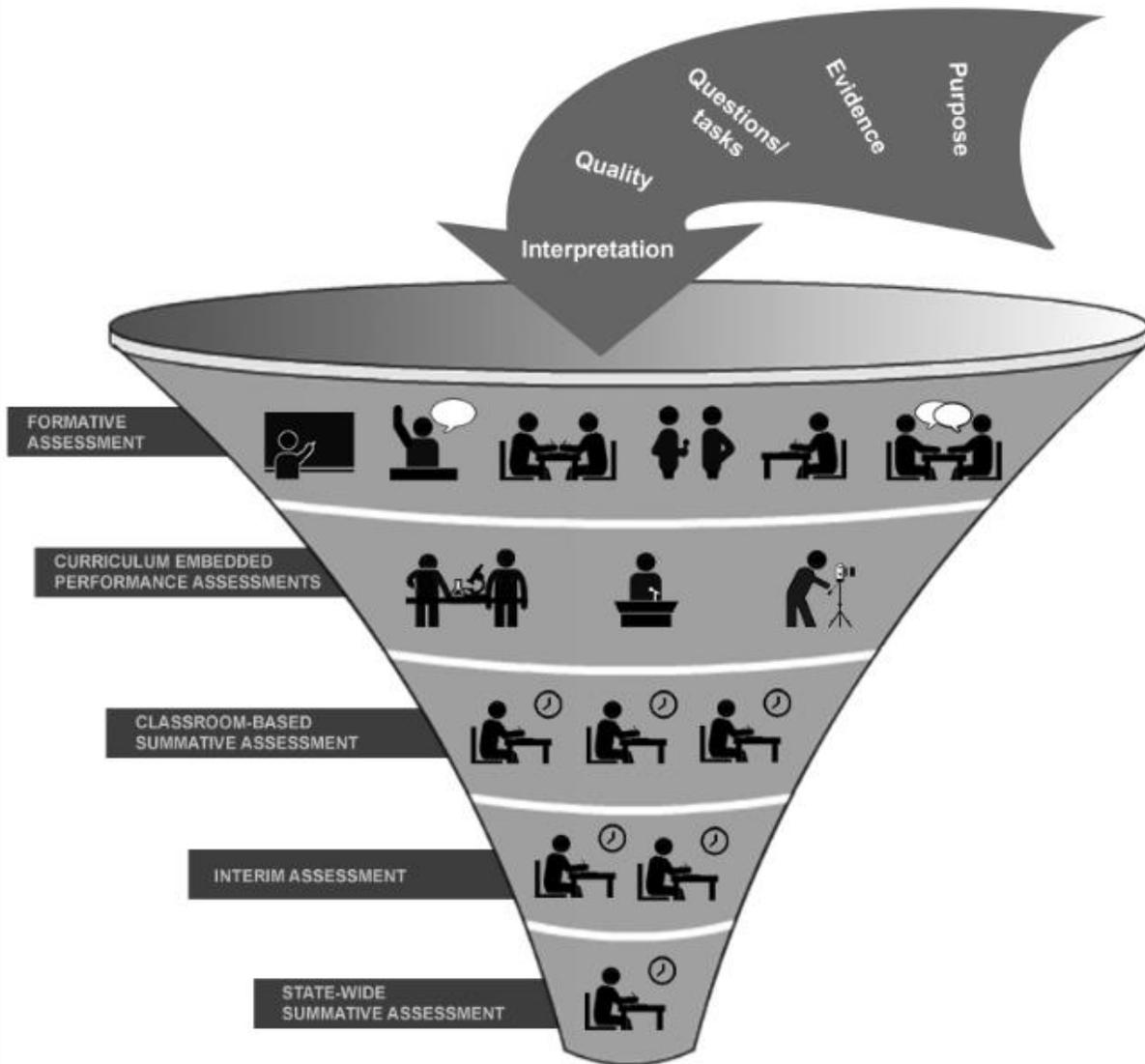


Figure 2. Sequencing teacher learning from the top layer down. Copyright © 2017 Educational Testing Service. Reprinted with permission.