

Rethinking the Test Pile

Every Test Needs a Purpose



Education First, Achievement Network,
and the case for coherent assessment

A purpose becomes credible when people can act

Every scheduled test asks students for time and attention. Leaders should be able to explain what happens to the answers.

Education First's *Rethinking the Test Pile* studies make the current assessment collection visible. ANet's work helps us examine the organizational practices around using evidence. Their contributions are complementary: one investigates the system; the other offers examples of how educators organize assessment within it.

The history connecting them reaches back decades. A useful assessment needs a defensible purpose, evidence suited to that purpose, and people equipped to respond. Audits begin this continuing responsibility.

Start with the next assessment-calendar or contract review.

Education First examines the whole test pile

The executive summary connects what districts administer, what states require, and what happens after an audit.



The entry point to three connected studies of district assessments, state policy, and assessment audits.

The district study asks what students take and how leaders use the results. The state scan examines the policies shaping those choices. The audit review looks at previous attempts to improve the system and the support needed afterward.

Rethinking the Test Pile: Executive Summary

Education First, 2026

Read together, they invite leaders to examine the assessment collection as a whole: where information overlaps, where a decision remains unanswered, and where requirements compete with the intended instructional use.

Test volume showed no significant link to achievement

The study provides a reason to investigate use, without identifying a universal number of tests to remove.



Rethinking the Test Pile: A National Study of K-8 Academic Assessments

Education First, 2026

Examines assessment volume and use across 67 districts in 38 states.

Within its purposive sample, the study found no statistically significant relationship between its volume measure and the achievement measures examined. The sample was not nationally representative, and this was not a causal test of reducing assessment.

For a district, the next step is to examine the measures already on its calendar: which decisions they support, whether their results are used, and what information another source already supplies.

Some choices begin outside the district

A local inventory also needs a clear account of what state policy requires.



Rethinking the Test Pile: A Scan of State Assessment Policies

Education First, 2026



Rethinking the Test Pile: State Assessment Audit and Assessment Literacy Efforts

Education First, 2026

The policy scan covers 27 states, with varying depth across policy areas. District-administered measures can still respond to state requirements.

The audit review finds limited documentation of sustained redesign alongside some lasting changes, including testing-time caps. Earlier audits provide useful lessons; continuing improvement also needs policy coordination and assessment literacy.

Separate local discretion from conflicts that require state action.

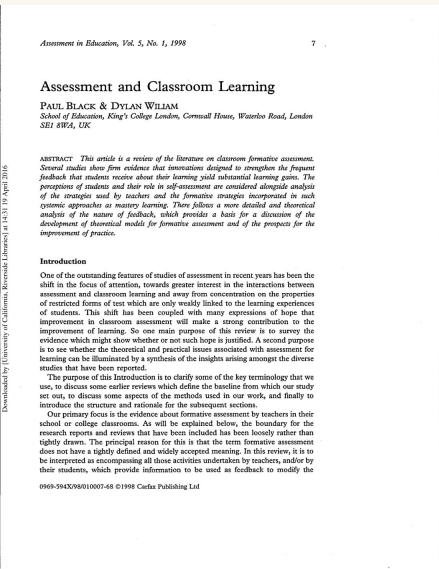
Useful feedback depends on a response

After the inventory comes a harder question: what makes retained evidence useful? The early literature begins with how teachers and learners respond.



Formative Assessment and the Design of Instructional Systems

D. R. Sadler, 1989



Assessment and Classroom Learning

Paul Black and Dylan Wiliam, 1998



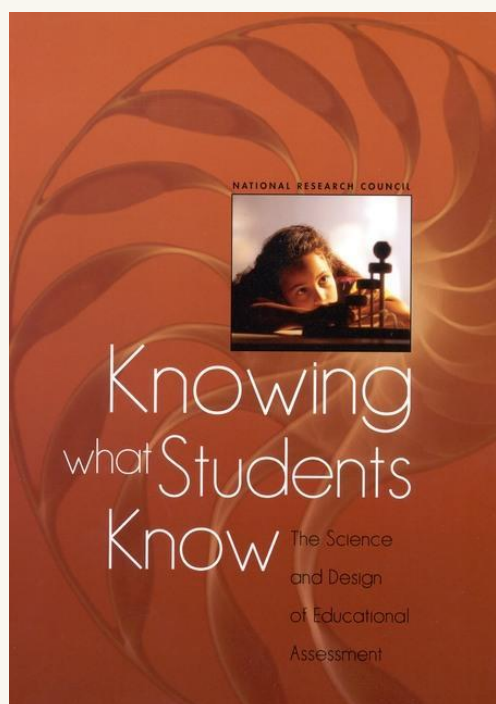
The Role of Assessment in a Learning Culture

Lorrie A. Shepard, 2000

These works direct attention to feedback, student participation, and the relationship between assessment and teaching. For leaders, they support a practical distinction: purchasing an interim test does not automatically create the classroom practices studied in formative assessment research.

Assessments need to work as a system

A district needs evidence that fits together around a shared vision for learning.



Knowing What Students Know: The Science and Design of Educational Assessment

National Research Council,
2001

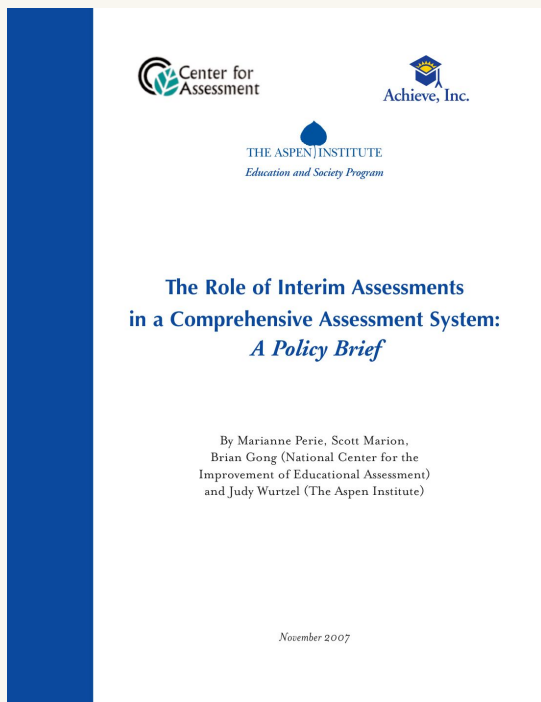
Makes a coherent account of learning the foundation for assessment design.

Knowing What Students Know called for systems that are comprehensive, coherent, and continuous. Classroom and larger-scale assessments should offer complementary evidence, connect to learning, and help show progress over time.

Coladarci's 2002 house-and-bricks analogy, cited by Chattergoon and Marion in 2016, captures the problem: collecting components does not establish how they work together. It's a pile of bricks rather than a well-built wall. A testing calendar can be complete while the reasons for its assessments remain disconnected.

Intended use comes before the purchase

Before buying a measure, a district needs to know who will use its results and why.



The Role of Interim Assessments in a Comprehensive Assessment System: A Policy Brief

Marianne Perie, Scott Marion,
Brian Gong, and Judy Wurtzel,
2007

Starts assessment selection with intended users, decisions, and support.

The brief asks what information is needed, who will use it, which actions should follow, and what professional support those actions require. Those answers guide design and reporting.

It distinguishes instructional, evaluative, and predictive uses. A measure designed for one purpose deserves scrutiny before leaders rely on it for another. Evidence for classroom formative practices cannot simply be inherited by a product described as formative.

Write the intended decision into the purchasing question.

A score needs a defensible interpretation

An intended use still requires a defensible interpretation of what students did.



Four Metaphors We Need to Understand Assessment

Robert J. Mislevy
Educational Testing Service

¹ Prepared for the Gordon Commission on the Future of Assessment in K-12 Education. I am grateful to John Behrens, Randy Bennett, Jim Pellegrino, and Lorrin Shepard for comments on an earlier version.

Examines assessment as practice, feedback, an evidentiary argument, and measurement.

Mislevy's four perspectives help leaders follow that reasoning. Measurement summarizes patterns. An evidentiary argument explains the conclusions those patterns support. Feedback concerns the response, while practice draws attention to the people and routines giving assessment meaning.

Four Metaphors We Need to Understand Assessment

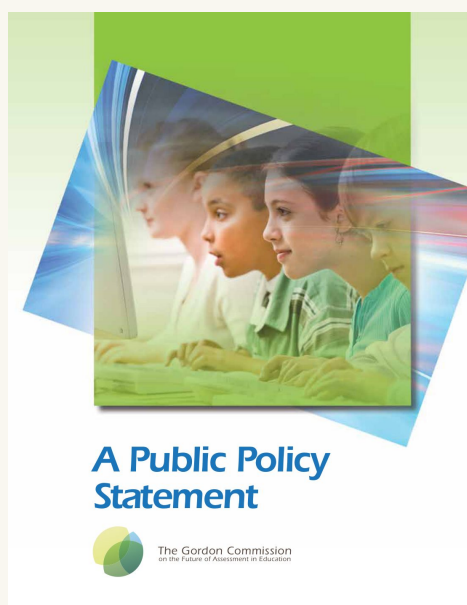
Robert J. Mislevy, 2013

Report first page

A district-level pattern can inform a resource decision. A teacher considering a particular student also needs evidence close to that student's work and curriculum. Moving results between users requires renewed judgment.

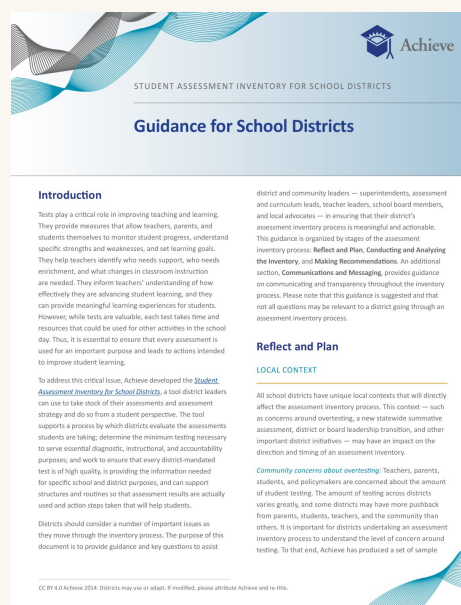
Different purposes belong in one system

Different purposes can coexist when leaders are explicit about their respective demands.



A Public Policy Statement

Gordon Commission on the Future of Assessment in Education, 2013



Student Assessment Inventory for School Districts: Guidance for School Districts

Achieve, 2015

The Gordon Commission called for a better balance while preserving accountability. Achieve's inventory guidance already asked districts to consider useful information and the routines for acting on it.

The historical task was broader than counting tests. The continuing question is how to sustain those commitments as requirements and priorities change.

The 2016 article makes system design concrete

Chattergoon and Marion make the connection between sound principles and daily decisions explicit.



Connects coherence, a theory of action, and efficiency.

Coherence asks whether assessment fits a model of learning and aligns with curriculum and instruction. A theory of action identifies users and explains how information should lead to a response. Efficiency examines redundant, unused, or untimely measures.

Not as Easy as It Sounds: Designing a Balanced Assessment System

Rajendra Chattergoon and
Scott Marion, 2016

Article first page, The State
Education Standard, NASBE

These criteria work together. Removing a test requires knowing what information is needed; retaining one requires explaining how its information reaches someone who can use it. One assessment need not do every job.

Review the system's connections as carefully as its individual tests.

Implementation crosses organizational boundaries

The principles remain demanding because responsibility is distributed across institutions and roles.



Examines the institutional barriers to putting sound assessment principles into practice.

The authors examine policy and political boundaries, commercial proliferation, weak connections to curriculum, and limited assessment literacy. Different offices can make defensible individual choices while leaving their combined effects unresolved.

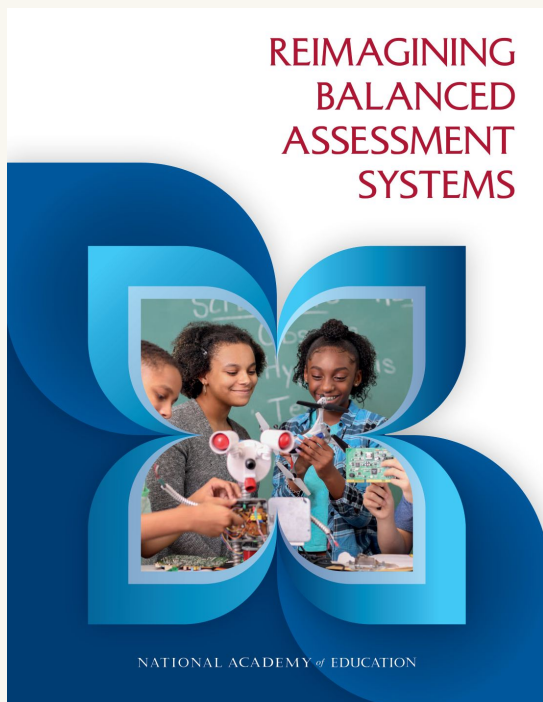
The Challenges and Opportunities of Balanced Systems of Assessment: A Policy Brief

S. Marion, J. Thompson, C. Evans, J. Martineau, and N. Dadey, 2019

Their analysis helps explain why a new tool or an isolated training session may leave the larger problem intact. District leaders need a way to coordinate assessment with curriculum and professional learning, including the state requirements they cannot change alone.

Balanced systems begin with learning

Two decades after the earlier system vision, implementation remains a central concern.



Reimagining Balanced Assessment Systems

Scott F. Marion, James W. Pellegrino, and Amy I. Berman, editors, 2024

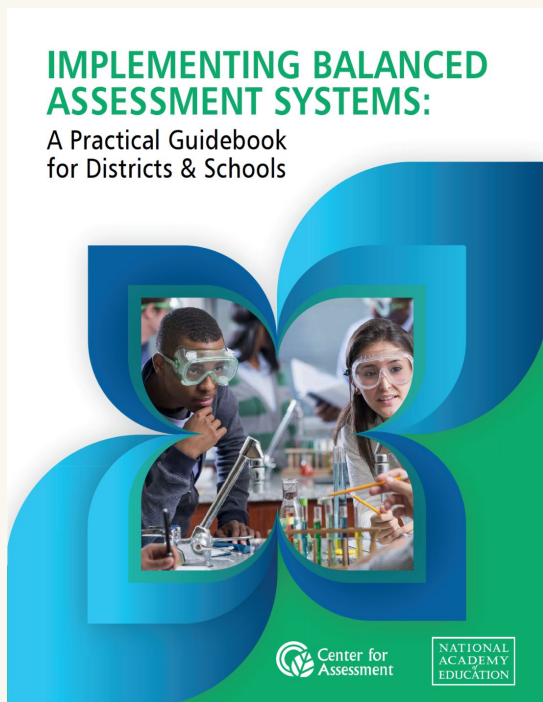
Revisits system design in service of ambitious and equitable teaching and learning.

This volume grounds assessment in ambitious and equitable classroom learning. Its chapters connect the design of evidence to professional learning and the responsibilities of districts and states.

Read Chapters 1–2 for the governing argument and implementation history, then Chapters 5–7 for the institutional work. The leadership question is concrete: what changes around teaching and curriculum would make the assessment system useful for the learning the district values?

The guide makes the design work practical

A shared ambition becomes useful when leaders can investigate how their present system works.



Implementing Balanced Assessment Systems: A Practical Guidebook for Districts and Schools

C. M. Evans, E. C. Wylie, E. Landl, and S. F. Marion, 2025

Helps districts examine efficiency, usefulness, and coherence, then manage change.

The guide examines excessive testing, redundancy, and unused results. It also asks whether assessment design fits its purpose and whether claimed instructional uses are plausible.

Its guidance extends into managing change so that improvement can survive staff turnover. This makes it a practical companion to the 2024 volume: leaders can identify a specific weakness, decide what to change, and organize the support needed to sustain the change.

ANet connects assessment design with instruction

The Missing Link brings the discussion to the relationship between assessment and instructional materials.



States ANet's current priorities for connecting assessment with instructional materials and decisions.

ANet's public overview focuses on evidence educators can use, a working connection with curriculum, and connected decisions across classrooms and district leadership.

These priorities offer design questions for leaders to examine locally. They describe ANet's approach. The discussion here draws on its public overview rather than an independent impact evaluation.

The Missing Link: Why Curriculum and Assessment Must Work Together

Achievement Network, 2025
Public overview of the white paper

Madison connects the audit with daily work

The district's assessment review became a plan for how people would work together.

VOLUME III | CHAPTER 7

From Evaluation to Impact: Transforming Assessment into a Tool for Learning

Michelle Odemwingie and Kimberly Cockrell
The Achievement Network, Ltd.

Abstract

This chapter examines the evolving role of assessment, moving beyond rigid measurements toward a more dynamic, learning-centered approach. Traditional assessment models often prioritize evaluation over instructional impact, but research and practice show that assessments can do more—guiding teaching, informing student learning, and strengthening instructional coherence.

The Achievement Network (ANet) works alongside schools to design assessments that emphasize transparency, alignment, and student agency. By integrating high-quality instructional materials with assessments that provide timely, meaningful feedback, ANet supports educators in making informed instructional decisions that drive student growth.

Key areas discussed include:

- **Transparency**, ensuring educators and students can understand purpose and design of assessments and act on results.
- **Coherence**, aligning curriculum, instruction, and assessment for a more structured learning experience.
- **Student-centered design**, fostering engagement and self-efficacy.

245

MADISON, WISCONSIN

ANet's 2021 case describes an inventory, recommendations to keep, cut, or modify measures, and teacher and family input. The 2025 chapter adds coordinated reporting and recurring interpretation in professional learning communities.

These are provider-authored implementation accounts. They make the work visible without establishing an independent achievement effect.

From Evaluation to Impact: Transforming Assessment into a Tool for Learning

Michelle Odemwingie and
Kimberly Cockrell, 2025

Chapter first page

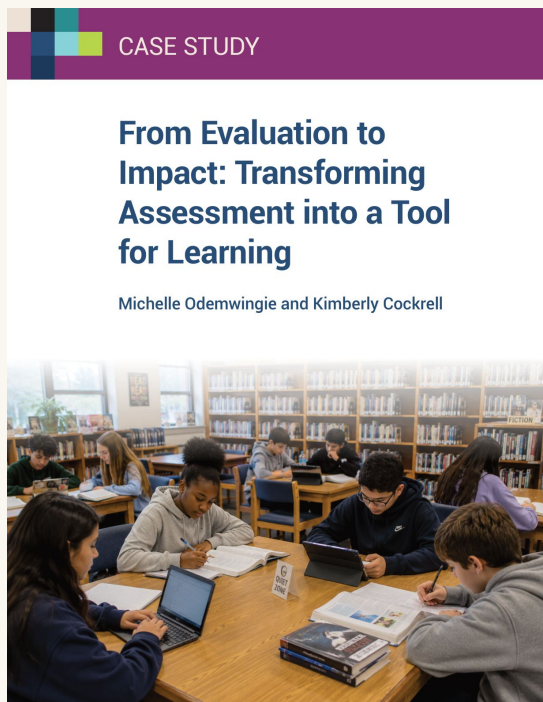
Case Study: Creating District Coherence Through Assessment Strategy

Achievement Network, December 18, 2021

Official case-study page

ANet's cases connect results with teaching

Carlsbad and Honey Dew make the work concrete: what educators do with results, and the support that response requires.



From Evaluation to Impact: Transforming Assessment into a Tool for Learning

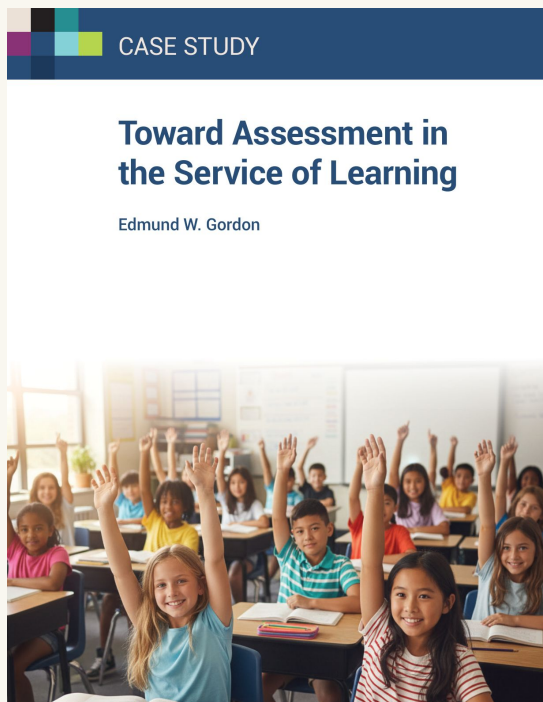
Michelle Odemwingie and
Kimberly Cockrell
2025 chapter
Case-study edition, May 2026

Practice accounts from Carlsbad and Honey Dew

In Carlsbad, leaders involved teachers in curriculum selection, used instructional walkthroughs alongside assessment results, and established an instructional leadership department. The authors describe how these arrangements supported shared expectations for teaching and identified where professional learning was needed. At Honey Dew, leaders and ANet coaches shifted toward more interactive professional learning to help teachers use results. The account describes David, a fifth grader struggling with fractions, whose teacher used assessment results to target instruction. These provider-authored examples describe practice without establishing an independent achievement effect.

Gordon asks assessment to develop learners' abilities

Gordon extends the question from using results to developing learners' abilities. His essay offers a vision for assessment design.



Toward Assessment in the Service of Learning

Edmund W. Gordon
November 2025

Connects assessment with learners' developing abilities.

Gordon recalls Haeussermann's work with children with neurological impairments. She focused on the conditions under which each child could succeed. Gordon develops that lesson into an argument for assessment that reveals how students think, respond to guidance, and learn through a task.

For leaders, the essay broadens the review of assessment purpose: consider what students can learn from the experience, including feedback and opportunities to try again. Gordon's conceptual argument invites leaders to consider the learner's experience when choosing evidence and support.

Different decisions need different evidence

Use these examples to ask whether the evidence fits the decision. They do not prescribe products.

Decision	Useful evidence and its limits	Possible response
Screening	A credible signal of possible risk or need. A screening result alone does not establish a diagnosis.	Arrange appropriate follow-up or further assessment.
Instruction	Student responses connected to current learning and curriculum. A broad score may leave the specific next step unclear.	Choose a teaching response and check subsequent work.
Public accountability and equity monitoring	Comparable results, considered with context and opportunity to learn. Group differences alone do not establish their causes.	Investigate patterns and review support or resources.

Choose evidence that fits the intended decision.

The next review makes purpose concrete

Begin with one district assessment proposed for adoption or due for renewal. Review the proposed or existing contract and its place on the calendar.

The board oversees the district's assessment choices; educators retain responsibility for classroom interpretation and instruction.

01 Name the decision and the response

Ask the team to name its purpose and explain why the evidence fits the decision. For a renewal, show how the results were last used. Identify the responsible user and the response that person has authority and support to make.

02 Make the choice and preserve needed evidence

Then decide whether to adopt, retain, change, or remove the measure. Record required elements separately from local choices. If a requirement creates a conflict, name the state-level action needed; if a measure is removed, check which decisions still need evidence.

Set a follow-up to examine what happened after the results arrived.

Leaders are responsible for what follows

Teachers need time to study responses, access to useful instructional options, and room to exercise judgment.

Education First gives leaders a reason to examine the whole assessment collection. The historical literature supplies durable design principles. Current guidance and ANet's accounts make the organizational work more visible.

The next step belongs to each system. Ask the superintendent to bring one adoption or renewal decision to the next planned review, with a clear account of the evidence, who will use it, and the support they need. Agree when to return to the decision and examine whether the response helped.

Every test needs a purpose. Leaders must make that purpose credible.

The featured studies and practice accounts

Rethinking the Test Pile: Executive Summary

Education First, 2026

Rethinking the Test Pile: A National Study of K–8 Academic Assessments

Education First, 2026

Rethinking the Test Pile: A Scan of State Assessment Policies

Education First, 2026

Rethinking the Test Pile: State Assessment Audit and Assessment Literacy Efforts

Education First, 2026

Reimagining Balanced Assessment Systems

Scott F. Marion, James W. Pellegrino, and Amy I. Berman, editors, 2024

Implementing Balanced Assessment Systems: A Practical Guidebook for Districts and Schools

C. M. Evans, E. C. Wylie, E. Landl, and S. F. Marion, 2025

The Missing Link: Why Curriculum and Assessment Must Work Together

Achievement Network, 2025; Public overview of the white paper

From Evaluation to Impact: Transforming Assessment into a Tool for Learning

Michelle Odemwingie and Kimberly Cockrell, 2025 chapter; May 2026 edition

Case Study: Creating District Coherence Through Assessment Strategy

Achievement Network, December 18, 2021

Toward Assessment in the Service of Learning

Edmund W. Gordon, November 2025

The historical reading

Formative Assessment and the Design of Instructional Systems

D. R. Sadler, 1989

Assessment and Classroom Learning

Paul Black and Dylan Wiliam, 1998

The Role of Assessment in a Learning Culture

Lorrie A. Shepard, 2000

Knowing What Students Know: The Science and Design of Educational Assessment

National Research Council, 2001

The Role of Interim Assessments in a Comprehensive Assessment System: A Policy Brief

Marianne Perie, Scott Marion, Brian Gong, and Judy Wurtzel, 2007

Four Metaphors We Need to Understand Assessment

Robert J. Mislevy, 2013

A Public Policy Statement

Gordon Commission on the Future of Assessment in Education, 2013

Student Assessment Inventory for School Districts: Guidance for School Districts

Achieve, 2015

Not as Easy as It Sounds: Designing a Balanced Assessment System

Rajendra Chattergoon and Scott Marion, 2016

The Challenges and Opportunities of Balanced Systems of Assessment: A Policy Brief

S. Marion, J. Thompson, C. Evans, J. Martineau, and N. Dadey, 2019

Coladarci, T. (2002). Is It a House ... or a Pile of Bricks? Important Features of a Local Assessment System. Cited through Chattergoon and Marion (2016), p. 6.