CASE STUDY

Mind Frames for Improving Educational Assessment

John Hattie, Stephen G. Sireci, and Eva L. Baker

Mind Frames for Improving Educational Assessment

John Hattie, Stephen G. Sireci, and Eva L. Baker

Abstract

Assessment in education has long prioritized accountability over meaningful interpretation for learning. This chapter calls for a shift toward assessment in the service of learning, emphasizing insights into student progress, learning strategies, emotions, engagement, and self-regulation rather than just achievement. To support this, educators must develop assessment-capable learners who can interpret and act on assessment results. The authors introduce 10 mind frames to enhance assessment, promoting diagnostic and predictive uses, clear success criteria, instructional alignment, and a classroom culture that embraces errors as learning opportunities. They also explore how technology and AI can make assessments more adaptive and personalized. By embedding assessment within teaching and learning, these mind frames transform it from a compliance tool into a driver of student growth and educational improvement. This chapter calls for changes in how we think about educational assessments. We believe such changes are needed if assessments will truly serve learners.

Shifting the Focus to Improvement

There is value in accountability models of assessment in informing policy makers of the impact of curriculum reforms, investment of resources, and needs for the future. However, classroom assessment's main functions—to inform and improve teaching and learning—are more important. The claim is that assessments can be crucial aids to inform and improve teaching and learning, help educators see "their impact," and accelerate learning when used for diagnostic and predictive purposes.

There needs to be interpretation relating to diagnosis (Where are the students now?), progress (How are they going?), informing next steps (Where should they go next?), and summatively, to ascertain who has attained the success criteria of the lesson(s). Primarily, the power of assessments comes from feedback to teachers about their impact: What did they teach well, and what not? Who did they teach well, and who they did not? And were the gains made appropriate for the time and resources invested? That is, what, who, and how much. Then, students are the greatest beneficiaries when these three impact questions form a primary assessment function.

Being clear about the diagnosis (including prior performance and current status) and desired success maximizes the power of assessment to know how to accelerate growth best. There is rarely one right way between these two points, and here is where the art of teaching is so critical, especially to evaluate continually the impact of the choice of teaching, the concepts and misconceptions, learning from the errors that are made (they need to be teaching opportunities, not embarrassments, Hattie, 2023). Also, differentiation fundamentally means different ways and times to attain success. Optimal instruction relies on understanding where a student is on a learning progression, the ability to probe for understanding and misunderstandings, and the provision of guidance toward the next steps. Success involves the 'knowing that" (content and facts), the 'knowing how' (the patterns and deeper conceptual understanding), and 'knowing with' (transfer to near and far situations) (Hattie, 2023; Roland, 1968; Ryle, 1945).

This shift in focus aims to develop student assessment capabilities to drive their own learning (Fisher et al., 2023). That is, to teach students how to interpret the results of any assessment such that they can (a) learn to be involved in making decisions about Where to next?, (b) understand more what they know and can do and what they still need to learn to attain success; and (c) fundamentally hear, understand and act upon the interpretations their teachers and they make to improve the rate and depth of learning.

In the remainder of this chapter, we propose 10 mind frames to inform thinking about assessment so the power of diagnosis and interpretation can be realized, and assessment can truly be used to accelerate learning.

The Ten Mind Frames for Assessment

Mind frames are ways of thinking related to the cognitive patterns, perspectives, interpretations, evaluations, and mental models we use to interpret our world. They influence how we perceive information, make decisions, solve problems, and navigate with others. In schools this thinking is the precursor to choosing high-impact strategies, ensuring the fidelity of implementation, and evaluating if there have been important impacts on students from the teacher's delivery of this instruction. It is the teacher's thinking that leads to the choice of interventions, devising and explaining the learning intentions and success criteria, knowing when a student is successful in attaining those intentions or not, having sufficient understanding of the students' understanding that they bring to the task, and sufficient knowledge about the content to provide meaningful and challenging experiences in various progressive pathways to success in learning.

Mind Frame #1: Assessment in schools needs to consider both progress and achievement, learning emotions, and strategies.

When we consider "tests" in schools, the dominant focus is achievement, and all too rarely are there measures of progress, emotions, learning, climate, striving, or engagement. A fundamental thesis is that educators need to start with high-trust climates, teach learning strategies, and then focus on their students' progress to higher achievement. We should not start with achievement, as this distorts the conversations to privilege those who begin well above average, and enhanced achievement (no matter where the student starts) is supposed to be an outcome of schooling.

Too many believe a "high achieving school" is necessarily a "great school." If the students start above average, many of these high achieving schools can support 'cruising'—that is, adding no value. We need both high achievement and high progress (See Figure 1). Any scores from classroom assessments administered over time can be plotted (and effect size and other measures of progress determined); students thus become their own baselines, and all can see who made progress or not. The progression, more than the achievement, is critical for successful schools and learning. Positive emotions, such as happiness, curiosity, and excitement, can enhance learning by increasing motivation, attention, and memory.

This first (and most critical) mindset aims to best inform educators about their impact on students' learning experiences so they can make better diagnostic interpretations leading to more effective instruction and experiences to accelerate students towards the success criteria. The message is to not only look at achievement, but also the causes and correlates of achievement, such as the high trust, high expectations and inviting climate, the choice of learning strategies, the emotions that speed or impede the experience of learning, the consequential levels of engagement, skills in working alone or in groups, and the motivations to want to learn more and deeper about the topics we desire them to learn about.

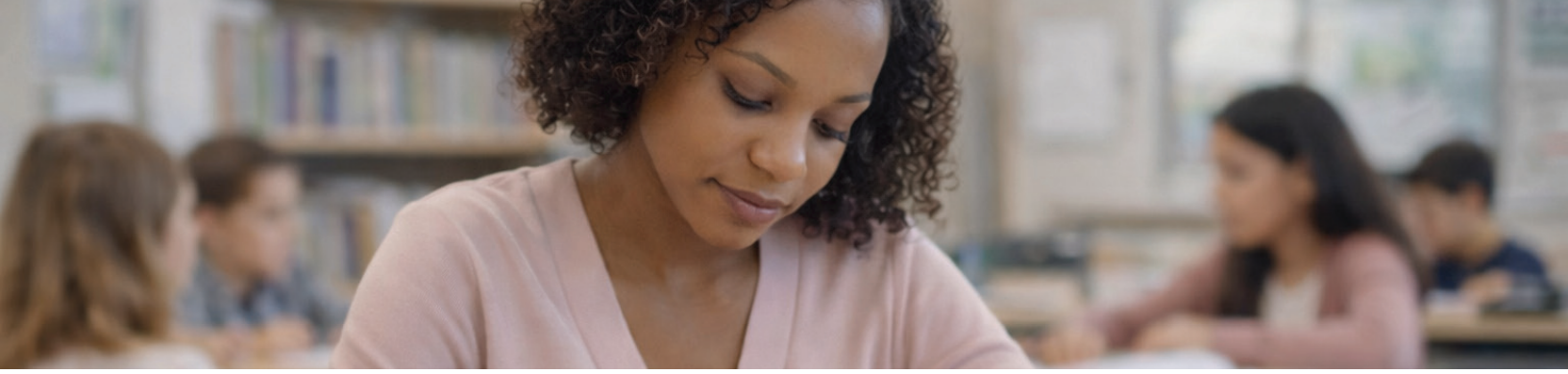
Mind Frame #2: Develop students' assessment capabilities so they can interpret the feedback and 'where to next' from assessments.

This mind frame emphasizes that we need to ensure students can (a) interpret the assessment results correctly and (b) make some consequential actions, decisions, or thinking that informs their next steps in learning. Sadler (1989, p. 143) noted "it is insufficient for students to rely upon evaluative judgments made by the teacher"; thus, requiring students' critical engagement in discerning the quality of their work and the criteria and standards against which their work is being judged (see also Baird, 2014).

Frey et al. (2019) argued that developing assessment capable learners leads to students becoming more aware of their current level of understanding in a learning area, more keen to understand their learning path and have renewed confidence to take on the challenge, better at selecting tools and resources to guide their learning, more ready to seek feedback and recognize that errors are opportunities to learn, more able to monitor their own progress and adjust course as needed, and recognize what they are learning and can teach others (see also Fisher et al., 2015). Thus, making students assessment-capable is perhaps one of the best ways to transform assessments into tools of learning.

To enable students to make these interpretations, teachers must model ways of using assessment information that helps students to meet their learning goals. In this way, students learn about: setting and clarifying challenging learning goals; how to access, interpret, and use evidence; understand the dimensions of engagement that lead to better outcomes; and engage in evaluative thinking about whether the work is good enough, meets the success criteria, and where to make the next learning moves.

Students can be empowered and have greater agency in their learning if they understand assessment results and how they lead to next actions. Teachers can help students acquire this understanding by modeling assessment information can help students meet their learning goals



Mind Frame #3: Formative, summative, and ascriptive evaluation refers to the timing and nature of interpretations (and NOT anything to do with kinds of assessment).

Scriven (1967) introduced the concepts of formative and summative evaluation, which are not intrinsically different types of evaluation but have different purposes. Formative evaluation is designed, done, and delivered to make improvements to the evaluation, and summative evaluation is done for, or by any decision-makers who need evaluative conclusions for any reason *other than* conceptual development. So, the distinction is that the purpose is formative during and summative at a key milestone: when the cook tastes the soup it is formative, when the guests taste the soup it is summative (Stake, cited in Miller et al., 2016). Both are important, both need to be appropriately rigorous, neither is more worthwhile than the others, both depend on the quality of interpretations, both require judgments.

One of the major fallacies is that there are concepts such as formative and summative testing. Give us any test, and we can make formative or summative interpretations. A test is neither formative nor summative; it depends on when an interpretation is made for improvement or the end of an intervention.

It is time to revert to Scriven's original claim: formative and summative evaluation—refers to evaluative thinking and evaluative decision-making when reviewing the outcomes of teaching, learning, and assessment opportunities (Clinton & Hattie, 2024). The quality and defense of the interpretations become critical (and not the test, sui generis).

We need to abandon notions of formative and summative tests because those terms are misnomers. Educational assessments can give us formative and summative information, and that it is the timing and nature of the interpretations that make them valuable.

Mind Frame #4: There are at least three levels of knowing: Knowing that (concepts, ideas, facts, surface), knowing how (relations, deeper conceptual), and knowing with (transfer, pattern recognition)—and assessment may need to measure each separately.

Contrary to many test developers' claims, a majority of the items on standardized assessments in the USA can be answered by simply knowing lots, especially in knowledge-rich subjects like science and history (Koretz, 2017). However, most theories of schooling ask for more than knowledge-rich graduates. There is also a clamor for students who can see patterns across ideas, have deeper conceptual knowing, and can transfer ideas from one to another situation or problem. The need is not either facts or deep, but facts and deep—depending on the nature of the problem.

A powerful taxonomy is SOLO (Structure of Observed Learning Outcome) taxonomy. The SOLO taxonomy provides a valuable framework for understanding and assessing learning outcomes as it is based on a model of cognitive complexity that allows educators to identify the level of understanding that students have reached and to design appropriate learning activities and assessments that challenge students to progress to higher levels of understanding (know where a student is performing and aim one step higher). Students can also use it to reflect on their learning and identify areas where they need to improve their understanding.

Clinton et al. (2021) reviewed Bloom, Depth of Knowledge, and SOLO and developed a model that brought all three together. Their four levels are categorized into two attributes: 'knowing that'/surface and 'knowing how'/deep thinking. The first major attribute of cognitive complexity relates to knowing that or surface thinking—the ideas, the factual, and content knowledge. This includes:

1. Factual knowledge recall and reproduction: for example, 'I know and can distinguish various parts of human hand anatomy.'
2. Conceptual knowledge involving basic application and skill: for example, 'I am able to apply Van Gogh's painting techniques to my drawing of the Sydney Opera House.'

The second major attribute is *knowing how* to think deeply or develop relations between ideas, extending to near or far new situations.

1. Strategic or relational thinking: For example, 'I understand that empowerment evaluation design principles have been utilized to design this evaluation compared to other evaluation models.'
2. Transfer: for example, 'I will be able to apply and adapt my methodological learning from my previous understanding of jazz principles to this new piece of music.' This can also be termed *knowing with.*

Clinton et al. (2021) argued that underlying these four levels is the notion of "evaluative thinking" or self-regulation–knowing when to be surface and when to be deep, and this skill invokes a particular kind of critical thinking and problem-solving.

Assessments need to be developed, scored, and reported at these four levels, or at least at the "Knowing that" compared to the "Knowing how and with" levels. This is not to be confused with difficulty at each level (factual, strategic, etc.). There can be increasing difficulty levels, but cognitive complexity moves down the depths from surface to deep to transfer.

Mind Frame # 5: We care about alignment between standards, success criteria, lessons, tasks, and assessments, and recognize the power of backward design (e.g., reports to tests, summative goals to influence formative directions, etc.).

When trying to understand the underlying reasons for the very discrepant effect sizes of many teaching methods, Hattie (2023) developed the 'Intentional Alignment model'. This model relates to the proportion of knowing that, how, and with (surface, deep, transfer) in the lessons, and then aligning these the notions of success (e.g., one success criteria for knowing that, and one for knowing how and with), the methods of teaching, the optimal strategies of learning, the cognitive requirements in the activities, and the assessment methods.

There are seven major parts of the Intentional alignment model:

1. Determining the learning intentions and success criteria, typically based on a curriculum, understanding progressions (where the students have been progressing, where they are now, and where they need to be), and the motivations and dispositions students bring to the class.
2. Cognitive task analysis of the knowing that, knowing how, and knowing with foci of the lessons (i.e., content, deeper understanding, and transfer skills and knowing).
3. Creating a climate and culture of the class to ensure students and teachers see errors or misconceptions as opportunities to learn.
4. Teaching methods that align with the cognitive complexity of the various components of the tasks, the required confidence to take on the challenges of this complexity, and efficiently and effectively improve students' progress towards the success criteria.
5. Ensuring students have appropriate and effective learning strategies to engage in the complexity of the tasks.
6. Choosing activities that align with the content's complexity levels, the deeper conceptual and relational thinking.
7. Using assessment methods that focus on the content ('knowing that'), the relational ('knowing how') and transfer ('knowing with') that inform teachers and students of their progress, success, and gaps to be then addressed.

These aspects of alignment are not only critical for using assessment to promote learning, but they are also essential for providing (a) evidence of the validity of these classroom assessments, and (b) professional development for teachers as they improve their teaching and formative evaluation processes. As the AERA et al. (2014) Standards stated, "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11). The curriculum and goals of instruction provide the theory to support test score interpretation and use, and confirmation of alignment provides the evidence to support such interpretation and use.

Alignment is needed across all stages of the education process, from goal and standard formulation to assessment, interpretation, and feedback. The Intentional alignment model provides a comprehensive way for building and evaluating such alignment.

Mind Frame #6: Provide scoring rubrics and success criteria near the beginning and not the end of lessons/courses to provide verisimilitude to any assessment.

It is often the case in schools when students are told to "do the work" with little to no understanding of what is 'good enough,' what the criteria of success look like, or now know the evaluative decisions to be made to claim success. However, when students can explain what they are learning, there is a 73% (from 37% to 64%) improvement in enjoyment and engagement in their learning.

There must be appropriately challenging expectations embedded in the success criteria. The Goldilocks principle of challenge is optimal: not too hard, not too easy, not too boring. Hence, it is important to specifically know, plan for, and develop confidence to take on appropriate challenges (self-efficacy), and have high levels of trust that failure is the best friend of learning in this class.

Scoring rubrics are more effective when they are provided at the outset rather than at the end of activities, and even better when students are involved in creating them (Becker, 2016), when they relate to lessons over time (and not short term(Andrade & Valtcheva, 2009). By providing students with the rubrics with which their work will be evaluated, they will approach the tasks more confidently, and thus enhancing their learning.

Mind Frame #7: We create a climate of welcoming errors and not knowing; thus, a major role of assessment is diagnosis and discovering where students are currently knowing and not understanding.

We do not come to class to learn what we know and understand, so the starting point is 'not perfection.' Sadly, so many learn quickly that classrooms privilege those who know and can do, eagerly respond to teacher questions, and know how to do the work. Too much assessment is making a status report on where 'they are now,' which is valuable, but the starting point to 'where to next' is the essence of learning.

Feedback thrives on errors, and undertaking challenges leads to errors. Therefore, the climate of the class must welcome errors, teach students how to engage in error detection, have teachers positively engage in error repair, and failure needs to be a learner's best friend. Errors can serve as important feedback information, indicating where the student's thinking and knowledge are not yet developed.

Many learning models explicitly include attention to errors and failure. Productive failure invites students to solve, usually ill-structured or complex, problems before instruction to evaluate their disequilibrium, try and invent multiple solutions, and realize what they need to learn from the subsequent instruction (Kapur, 2024). Thus, there needs to be a climate for receiving, welcoming, and learning from test information.

Classrooms should embrace a culture where mistakes are welcomed and discussed. In such environments, the diagnostic role of assessments can help students discover where they are and what to do next improve their learning

Mind Frame #8: We act on the belief that a major purpose of assessment in schools is to inform teachers of their impact.

This is the major message of the Visible Learning work—we care less about how teachers teach and assess but care more about the impact of their teaching and assessment interpretations. Switching from the act of teaching and testing to the *impact* changes the debate, makes it easier to recognize expertise, and puts the focus on the dependability of the interpretations and consequential actions. This focus on impact means interpretation of multiple sources of evidence, including test scores, is critical. The key is triangulating this evidence from tests, teacher noticing, student voices about their learning, and artifacts of student work. Educators can use assessment to assess their impact and the impact of the different teaching tools and instructional practices they use.

Mind Frame #9: Many technologies (especially large language models) can make developing and scoring assessments more efficient and lead to making assessment interpretations more defensible.

The most remarkable change in our lives is the advent of large language models such as ChatGPT, Gemini, Claude, etc. While the usual claims about evaluating the dependability and validity remain (and become perhaps even more important), these AI models can more readily 'write' items to specifications, offer more opportunities to adapt tests on the fly, score open-ended assessments, and create more tailored interpretations from the assessments. Maybe the over-dependence on multiple-choice and closed items that have dominated many assessment systems (as they are cheaper and easier to score) will be reduced, and items that map the processes of learning, the deeper ideas, and the construction (rather than recognition) of answers will take their rightful place within tests.

Sireci et al. (2024) proposed a new model using many of the advances from AI tools called Design-In-Real-Time (DIRTy) assessment, which reflects the progressive evolution in testing from a single test to an adaptive test, to an adaptive assessment system. It involves: (a) assessment building blocks (individual items or "assessment task modules" (ATMs) that are linked to multiple content standards and skill domains), (b) gathering information on test takers' characteristics and preferences and using this information to improve their testing experience, and (c) selecting, modifying, and compiling items or ATMs to create a personalized test that best meets the needs of the testing purpose and the individual test taker.

What is new in DIRTy assessment is tailoring the test to multiple, personal factors, and delaying test specifications until the interaction of a test taker and a testing purpose occurs. A major advantage is that by using the assessments to enable test takers to solve problems, the assessment becomes a vehicle on route to solving the problem, and hence has the potential to promote learning (Gordon, 2020). DIRTy assessments can also be more culturally responsive. For example, students can have choice in which reading passages or other stimuli they respond to, and the choices can reflect the communities in which test takers live.

Technology has much to offer education, and with respect to assessments that serve learners, we can use technology to develop interactive assessments that are optimal for each learner, rather than a single assessment optimized for most learners. In addition to developing and delivering tests, technology can provide real-time feedback from assessments, and engage learners in understanding their assessment performance and what to do next. Technology has the potential to allow assessments to foster engagement in the testing process, and to be fully aligned with and integrated into instruction.

Mind Frame #10: Transform the purposes of assessment to move away from an overreliance on accountability to more continuous assessments used in learners' acquisition of understanding, motivation for learning, collaboration, and problem-posing settings.

We need to reorient our practices to value students more than a score scale. Most classroom assessments rarely move past estimates of achievement, and few consider the students' strategies of learning, their emotions before, during, and after a lesson(s), and so often, the skills to work alone or in groups are considered. There needs to be a transformation of classroom assessment purpose from annual, time-controlled accountability assessments to more continuous assessments used in learners' acquisition of understanding, motivation for learning, collaboration, and deep application of knowledge in problem-solving, communication, and authentic settings.

Such student-targeted assessments will have more validity to understand what students know and where they need to go next. We do not need to stop instruction to assess students for accountability purposes; we need assessments as part of instruction to guide it. If we solve the problems of developing, administering, and interpreting valid assessments supporting student learning, aggregating the results from those assessments for accountability will be relatively easy. It is time for assessments to be supportive, not disruptive. It is time to make learners full partners in the assessment process so the results benefit them and place their needs above those of the policy makers.

Concluding Comments

We presented 10 mind frames for teachers, students, test developers, and all others involved in educational assessment that will move educational assessments from the measurement "of learning" to "supporting learning." Baker (2018) has argued that it is well nigh that we "connect assessments systematically to instruction.

Assessment should be organized as part of the instructional process rather than be disruptive to teaching and learning. Including assessment in instruction requires considering when student evidence of learning is needed, what form that evidence should take, and how the evidence will inform future learning and instruction. Without a doubt, the focus on improvement through design and teacher and student interpretations is essential, as are the moral and desired modifications to practice that support varied learners where they are.

Scaling-up issues aside, we continue to think that the mind frame shift is likely the most important catalyst for accelerating student learning at a broad level for all students. As the mind frames in this chapter illustrate, by engaging and empowering students in their learning, by making them feel comfortable with their mistakes and allowing them the freedom to explore them, and by thinking about how assessments can be developed and used differently, we will have progressed beyond 19th and 20th-century ways of thinking about tests, to using them to support student learning.

References

Absolum, M., Flockton, L., Hattie, J. A. C., Hipkins, R., Reid, I (2009). *Directions for Assessment in New Zealand: Developing students' assessment capabilities.* Ministry of Education, Wellington, NZ. http://assessment.tki.org.nz/Assessment-in-the-classroom/Directions-for-assessment-in-New-Zealand-DANZ-report

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into practice*, *48*(1), 12–19.

Baird, J. A. (2014). Teachers' views on assessment practices. *Assessment in Education: Principles, Policy & Practice*, *21*(4), 361–364.

Baker, E. L. (2004). *Principles for Scaling Up: Choosing, Measuring Effects, and Promoting the Widespread Use of Educational Innovation.* CSE Report 634. Center for Research on Evaluation Standards and Student Testing CRESST.

Baker, E. (2018). Design for assessment change. *European Journal of Education, 53*(2), 138–140.

Baker, E. L., & O'Neil, H. F. (2020). The assessment landscape in the United States: From then to the future. *Monitoring student achievement in the 21st-century: European policy perspectives and assessment strategies*, 51–61.

Bandura, A. (1986). *Social foundations of thought and action.* Englewood Cliffs.

Bandura, A. (1993). *Perceived self-efficacy in cognitive development and functioning. Educational Psychologist, 28*(2), 117–148.

Becker, A. (2016). Student-generated scoring rubrics: Examining their formative value for improving ESL students' writing performance. *Assessing Writing*, *29*, 15–24.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome).* Academic Press.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90.

Blannin, J., Hattie, J. A. C., Wood, C., & Stubbs, P. (in review). Informing Professional Learning Interventions with Evidence-Based Analysis of Student Feedback: Implications for Software Use and Learning Clarity

Bloom, B. S. (1971). *Handbook on formative and summative evaluation of student learning.* McGraw-Hill.

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, Handbook I: The cognitive domain.* David McKay Co.

Booth, N. (2017, July 9). *What is formative assessment, why hasn't it worked in schools, and how can we make it better in the classroom?* Impact. https://my.chartered.college/impact_article/what-is-formative-assessment-why-hasnt-it-worked-in-schools-and-how-can-we-make-it-better-in-the-classroom/

Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*, *36*(3), 375–388.

Buzick, H. M., Casabianca, J., & Gholson, M. L. (2023). Personalizing large-scale assessment in practice. *Educational Measurement: Issues and Practice*, https://doi.org/10.1111/emip.12551

Clinton, J. M., & Hattie, J. A. C. (2021). *Cognitive complexity of evaluator competencies.* Evaluation and Program Planning, 89, 1–8.

Clinton, J., & Hattie, J. (2024). Revisiting and Expanding Scriven's Fallacies About Formative and Summative Evaluation. *Journal of MultiDisciplinary Evaluation*, *20*(47), 13–23.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics, 132*(4), 1593–1640.

Fisher, D., Frey, N., Ortega, S., & Hattie, J. A. C. (2023). *Teaching students to drive their learning: A playbook on engagement and self-regulation.* Corwin.

Francis, E. M. (2022). *Deconstructing depth of knowledge.* Solution Tree.

Frey, N., Hattie, J. A. C., &, Fisher, D. (2018). *Developing Assessment Capable Learners*. Thousand Oaks, Corwin.

Gordon, E. W. (2020). Toward assessment in the service of learning. *Educational Measurement: Issues and Practice*, *39*(3), 72–78.

Gordon, E. W., & Rajagopalan, K. (2015). *The testing and learning revolution.* Jossey-Bass.

Hattie, J. A. C. (2014). The Last of the 20th-Century Test Standards. *Educational Measurement: Issues & Practice*, *33*(4), 34–35.

Hattie, J. A. C. (2018). *Implementing, scaling up, and valuing expertise to develop worthwhile outcomes in schools*. William Walker Oration, Presented at the Annual Conference of the Australian Council for Educational Leaders, Sydney. ACEL Monograph #58. http://www.acel.org.au/acel/ACELWEB/Publications/ACEL_Monograph.aspx

Hattie, J. A. C. (2023). *Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement*. Routledge.

Hattie, J. A. C., Brown, G., Ward, L., Irving, E., & Keegan, P. (2006). Formative evaluation of an educational technology innovation: Developer's insights into assessment tools for teaching and learning. *Journal of Multidisciplinary Evaluation, 5*(3), 1–54.

Hattie, J. A. C., Clarke, S., Fisher, D., & Frey, N. (2021). *Collective student efficacy.* Corwin.

Hattie, J. A. C., O'Leary, T., Hattie, K., & Donoghue, G. (2025). *Great Learners by Design.* Corwin.

Hattie, J. A. C., & Purdie, N. (1998). The Solo model: Addressing fundamental measurement issues. In Dart, B., & Boulton-Lewis, G. M. (Eds.), *Teaching and learning in higher education.* Camberwell, Vic, Australian Council of Educational Research.

Hess, K. (2006). *Exploring cognitive demand in instruction and assessment.* National Center for the Improvement of Educational Assessment, Dover NH.

Kapur, M. (2008). Productive failure. *Cognition and instruction*, *26*(3), 379–424.

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science*, *40*, 651–672.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299.

Kapur, M. (2024). *Productive Failure: Unlocking Deeper Learning Through the Science of Failing.* Jossey-Bass

Koretz, D. (2017). *The testing charade: Pretending to make schools better.* University of Chicago Press.

Law, N., Alexander-Hollins, Smith, D., & Hattie, J. A. C. (2024). *10 Mindframes for the culture and climate of schools: Equity, Identities, and Belonging.* Corwin.

Lee, J., LeBaron Wallace, T., & Alkin, M. (2007). Using problem-based learning to train evaluators. *American Journal of Evaluation*, *28*(4), 536–545.

Linden, W. J. (2005). *Linear models for optimal test design*. Springer Science+ Business Media, Incorporated.

Lord, A. Theory of test scores.-, 1952. *Psychometric Monograph*, (7).

Luecht, R. M., & Hambleton, R. K. (2021). Item response theory: A historical perspective and brief introduction to applications. In *The History of Educational Measurement* (pp. 232–262). Routledge.

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessments, and instruction, *Review of Educational Research 4*, 1332–1361.

Miller, R. L., King, J. A., Mark, M. M., & Caracelli, V. (2016). The oral history of evaluation: The professional development of Robert Stake. *American Journal of Evaluation*, *37*(2), 287–294.

Moeller, J., Brackett, M. A., Ivcevic, Z., & White, A. E. (2020). High school students' feelings: Discoveries from a large national survey and an experience sampling study. *Learning and Instruction*, *66*, 101301.

Oberauer, K. (2010). Declarative and procedural working memory: Common principles, common capacity limits?. *Psychologica Belgica*, *50*(3–4), 277–308.

Piaget, J. (1952). *The origins of intelligence*. New York: International University Press.

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review*, *31*, 339–359.

Roland, J. (1958). On "knowing how" and "knowing that". *The Philosophical Review, 67*(3), 379–388.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67.

Ryle, G. (1945, January). Knowing how and knowing that: The presidential address. In *Proceedings of the Aristotelian society* (Vol. 46, pp. 1–16). Aristotelian Society, Wiley.

Ryle, G. (1949). *The concept of mind*. Penguin Books.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144.

Sagone, E., & De Caroli, M. E. (2014). Relationships between psychological well-being and resilience in middle and late adolescents. *Procedia-social and behavioral sciences*, *141*, 881–887.

Schellings, G. (2011). Applying learning strategy questionnaires: Problems and possibilities. *Metacognition and Learning*, *6*, 91–109.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler (ed.), *Perspective of Curriculum Evaluation*, American Educational Research Association (AERA), Monograph of Curriculum Evaluation, No. 1., Rand McNally.

Scriven, M. (1993). The nature of evaluation. *New Directions for Program Evaluation*, *58*, 5.

Sireci, S. G. (2020). "De-"Constructing" Test Validation," *Chinese/English Journal of Educational Measurement and Evaluation*, 1. 教育测量与评估双语季刊:
https://www.ce-jeme.org/journal/vol1/iss1/3
https://doi.org/10.59863/CKHH8837

Sireci, S. G. (2020). Standardization and understandardization in educational assessment. *Educational Measurement: Issues and Practice, 39*(3), 100–105. https://doi.org/10.1111/emip.12377

Sireci, S. G. (2021). Valuing educational measurement. *Educational Measurement: Issues and Practice, 40*(1), 7-16. https://doi.org/10.1111/emip.1241

Sireci, S. G., Suárez-Alvárez, J., Zenisky, A. L., & Oliveri, M. E. (2024). Evolving educational testing to meet students' needs: Design-in-real-time assessment. *Educational Measurement: Issues and Practice*, *43*(4), 112–118.

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 26,* 347–376.

Surma, T., Vanhees, C., Wils, M., Nijlunsing, J., Crato, N., Hattie, J., Muijs, D., Rata, E., Wiliam, D., & Kirschner, P. A. (2025). *Developing Curriculum for Deep Thinking: The Knowledge Revival.* Springer

Thissen, D., & Steinberg, L. (2020). An intellectual history of parametric item response theory models in the twentieth century. *Chinese/English Journal of Educational Measurement and Evaluation*| 教育测量与评估双语期刊, *1*(1), 5.

Thurstone, L. L. (1925). *The fundamentals of statistics* (Vol. 4). Macmillan.

Traub, R. E., & Wolfe, R. G. (1981). Chapter 8: Latent Trait Theories and the Assessment of Educational Achievement. *Review of Research in Education*, *9*(1), 377–435.

Tyler, R. W. (1958). Curriculum organization. *Teachers College Record*, *59*(11), 105–125.

Vo, A. T., Schreiber, J. S., & Martin, A. (2018). Toward a conceptual understanding of evaluative thinking. *New Directions for Evaluation*, *2018*(158), 29–47.

Webb, N. (1997). *Criteria for alignment of expectations and assessments on mathematics and science education.* Research Monograph Number 6. CCSSO.

Webb, N. L. (2002). Depth-of-knowledge levels for four content areas. *Language Arts*, *28*(March), 1–9.

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, *20*(1), 7–25.

Wyatt-Smith, C., & Adie, L. (2021). The development of students' evaluative expertise: Enabling conditions for integrating criteria into pedagogic practice. *Journal of Curriculum Studies*, *53*(4), 399–419.

Zenisky, A. L., O'Donnell, F., & Hambleton, R. K. (in press). Reporting scores and other results. In L. L. Cook & M. J. Pitoniak (Eds.), *Educational measurement* (5th edition). Oxford University Press.

Study Group

About the authors

John Hattie is Emeritus Laureate Professor at the Melbourne Graduate School of Education at the University of Melbourne, Chief Academic Advisor for Corwin, i-Ready Technical Advisor, and co-director of the Hattie Family Foundation. His career was as a measurement and statistics researcher and teacher, and his more recent research, better known as Visible Learning, is a culmination of nearly 30 years synthesizing more than 2,500 meta-analyses comprising more than 140,000 studies involving over 300 million students around the world.

Stephen G. Sireci, Ph.D., is Distinguished Professor and Executive Director of the Center for Educational Assessment in the College of Education, University of Massachusetts Amherst. He earned his Ph.D. in psychometrics from Fordham University and his master and bachelor degrees in psychology from Loyola College Maryland. Before UMass, he was Senior Psychometrician at GED Testing Service, Psychometrician for the CPA Exam and Research Supervisor of Testing for the Newark NJ Board of Education. He is known for his research in validity and fairness of educational tests, and for innovations in test development.

Eva L. Baker is a Distinguished Professor at UCLA and founding Director of the Center for Research on Evaluation, Standards and Student Testing, (CRESST). She is widely published in the areas of learning-based assessments, technology, and policy. She served as Chair of the Board on Testing and Assessment, National Research Council, and Co-Chair of the 1999 Standards for Educational and Psychological Testing. Baker served as president of the World Education Research Association (WERA) and was president of the American Educational Research Association (AERA). A member of the National Academy of Education, she received AERA's Robert L. Linn Lecture and the E. F. Lindquist Award.

About the Study Group

The Study Group exists to advance the best of artificial intelligence, assessment, and data practice, technology, and policy; uncover future design needs and opportunities for educational systems; and generate recommendations to better meet the needs of students, families, and educators.

Date of Publication
February 2026

Citation
Hattie, J., Sireci, S. G., & Baker, E. L. (2025). Mind frames for improving educational assessment. In E. M. Tucker, E. Armour-Thomas, & E. W. Gordon (Eds.), *Handbook for assessment in the service of learning, Volume I: Foundations for assessment in the service of learning.* University of Massachusetts Amherst Libraries.