CASE STUDY

Innovating Assessment Design to Better Measure and Support Learning

Natalie Foster and Mario Piacentini

Innovating Assessment Design to Better Measure and Support Learning

Natalie Foster and Mario Piacentini

Abstract

Assessments should support deep learning as well as measure its outcomes. This piece proposes five design principles–following insights from research in the learning and cognitive sciences–for designing innovative assessments to measure and support the development of complex competencies.

These principles include:
1) using extended performance tasks;
2) accounting for students' prior knowledge;
3) providing opportunities for productive failure;
4) providing feedback and instructional support; and
5) designing "low floor, high ceiling" tasks.

Our fundamental argument driving these design principles is that assessments should not only measure student progress towards educational goals, but also model and provide insight into students' deep learning processes.

Adopting these design principles will enable educators to collect information about how well students can engage in complex thinking and problem-solving processes while reducing the current distance between assessment and learning.

Introduction

In recent years, researchers have argued for the need to innovate educational assessments to better measure and support the development of important skills (e.g. Foster & Piacentini, 2023; Kyllonen & Sevak, 2024; Schwartz & Arena, 2013). Such works respond to shifts in educational discourse and policy about what is important to teach and learn—so-called 21st-century competencies. Innovation is required across the entire assessment development process: from conceptual foundations (defining the components of what are often complex constructs and the authentic contexts in which they are engaged), to design considerations (how tasks are designed and delivered to test takers), measurement issues (how to generate, interpret and accumulate useful evidence about what students know and can do), and reporting options (how to clearly communicate the results to intended audiences, be they teachers, learners, administrators or policy makers). To achieve positive impact on educational outcomes, more coherent systems of assessment are also needed, increasing the alignment between formative and summative assessment (Darling-Hammond et al., 2013; Pellegrino, 2023).

One of the key arguments of our approach is that assessments should not only be useful for measuring student progress towards educational goals, but they should also model and provide insight into students' deep learning processes (Piacentini et al., 2023). In this way we can collect important data on complex thinking and problem-solving processes, and at the same time, we can reduce the current distance between assessment and the learning we want to promote in classrooms.

Arguments in support of innovating assessments

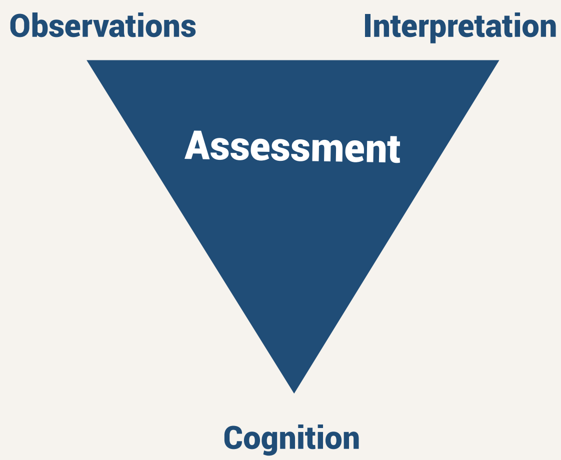
A fundamental conception about assessment is that it constitutes a process of reasoning from evidence, guided by theory about the critical aspects of knowledge and skill one is interested in measuring. This process of reasoning from evidence has been portrayed as a triad of three interconnected elements: the *assessment triangle* (Pellegrino et al., 2001). The vertices of the assessment triangle represent the three key elements underlying any coherent assessment: a model of student *cognition* and learning in the domain of the assessment; a set of assumptions and principles about the kinds of *observations* that will provide evidence of students' competencies; and an *interpretation* process for making sense of the evidence (See Figure 1).

FIGURE 1: *The assessment triangle*

Cognition
Theories, models & data about how students represent knowledge & develop competence in a domain of instruction and learning.

Observations
Tasks or situations that allow one to observe students' performance.

Interpretation
Methods for making sense of the evidence coming from students' performances.



Source: Pellegrino et al. (2001).

Measuring what matters

Broadly speaking, some distinct categories of competencies emerge across frameworks that set forth a vision for the future of education and the broad, transferable areas of knowledge and skills that students need, including cognitive (e.g., creative thinking, critical thinking, problem solving), interpersonal (e.g., communication, collaboration), metacognitive (e.g., self-regulated learning), intrapersonal (e.g., persistence, adaptability) and digital (e.g., media literacy, digital literacy) (Foster, 2023). Digital competencies are often a central component that intersect with other 21st-century competencies given the vast proliferation of ICTs and related advances (e.g., Artificial Intelligence) in both our personal and professional lives, in turn transforming our individual and collective capacities for communication, collaboration, searching for information and using knowledge.

Not only does this mean being able to measure the extent to which students master these important competencies in different contexts, but also designing assessments that can provide timely feedback to learners and educators that can benefit instructional practices. In other words, developing assessments that *support* deep learning as well as measure its outcomes.

Developing assessments that support deep learning as well as measure its outcomes is essential to preparing students for the complexities of the 21st century.

Several interconnected conceptual and practical challenges exist when designing assessments of 21st-century competencies, particularly in the context of large-scale assessment. These challenges include defining the assessment construct and learning progressions for complex competencies at different levels of proficiency, identifying the extent to which domain-specific knowledge supports performance and whether performance in an assessment can be generalized to other contexts, being able to identify and design tasks that elicit both outcomes and the processes that often define these competencies, and being able to generate interpretable evidence about students' ways of thinking and doing (see Foster, 2023, for an in-depth discussion of each).

Unlocking new advances in design and measurement technology

Advances in technology offer much potential to transform what we can measure, how we can make sense of test taker performance, and how assessment can relate to learning (Thornton, 2012; Shute & Kim, 2013; Timmis et al., 2016; Zhai et al., 2020; DiCerbo et al., 2017). Technology can introduce new forms of active, immersive and iterative performance-based tasks in interactive environments that make it possible to observe how test takers engage in authentic, open-ended learning and problem-solving activities. These can provide richer observations and potential evidence about students' thinking processes, behaviors and decision-making, as well as enable the measurement of dynamic skills beyond the capability of more traditional and static test items. These environments allow students to engage actively in the processes of making and doing, making it possible to track the strategies that test takers employ and the decisions they make as they work through complex tasks that evolve on their basis of their actions. Such student-centred tasks also provide opportunities for students to learn and develop skills as they make decisions and engage in iterative problem-solving.

New task modalities, problem types and affordances mean there are also new possibilities for response types—and in turn, new sources of potential evidence about test takers (Sabatini et al., 2023). Digital platforms can capture, time stamp and log student interactions with the test environment. This is especially transformative in the context of measuring 21st-century competencies, as the process by which an individual engages with an activity can be just as valuable for evaluating proficiency as their final product. These process data, when coupled with appropriate analytical models, can reveal how students engage with problems, the choices they make, and the strategies they do (or do not) implement. Patterns of behavior associated with different mastery levels can be identified, which can then be used to augment the precision of performance scores and provide diagnostic information to educators about students.

Taken together, these opportunities represent a powerful shift in assessment away from simple knowledge reproduction to knowledge-in-use—exactly the types of performances in which 21st-century competencies are engaged and developed. Measuring these constructs well necessarily requires assessments to be closer tied to the processes and contexts of learning and instruction—something that technology can facilitate through embedded assessment. Embedded assessment environments can integrate pedagogical affordances like scaffolding and feedback to explicitly support learning, merging summative and formative assessment and providing measures of the capacity of test takers to learn and transfer their learning to other tasks. By allowing for the real-time measurement of students' capacities as they engage in meaningful learning activities, technology holds the promise of creating new systems of evaluation where evidence on students' progress is collected in a continuous and unobtrusive way. In turn, educators can gain better insights into their students' learning processes, and assessment and learning are no longer explicitly separated activities.

Technology is transforming assessment from measuring what students know to understanding how they learn

Design principles for innovative assessments

While these opportunities are exciting, few extant assessment systems take full advantage of this potential to measure what matters. Traditional assessments (especially large-scale) have evolved to comply with practical constraints on testing time, cost efficiency and established measurement models.

In the rest of this chapter, we contend that it is possible to address this misalignment between current practice and promise in two ways: first, by taking stock of research that has investigated the mental structures that support the types of learning and problem solving encompassed by 21st-century competencies; and second—much more difficult—by creating an internally consistent system of teaching practices and assessments that reflect these research insights. In this system, deeper learning experiences prepare students for future learning, and innovative assessments measure how effectively students have engaged with these deeper learning experiences.

Insights about deeper learning processes and their implications for assessment

One of the main conclusions from research in the learning sciences is that learning is a socially situated process (Dumont et al., 2010; Darling-Hammond et al., 2019). That learning is mediated by socially constructed practices has clear implications for instruction: deeper learning occurs when students engage in activities that are realistic, complex, meaningful, and motivating, and when they can call upon the experience of knowledgeable others for guidance and support. Assessment experiences aimed at engaging and measuring deeper learning processes must therefore replicate these features.

Deep learning and skill mastery cannot be divorced from their contexts of practice; meaningful assessment must occur in context.

Some research has focused on contrasting "experts" (i.e., individuals that have constructed mental models in a given domain and who, through participation in key practices, have learned to apply these ideas to solve new problems) with "novices" (i.e., those who have not consolidated their basic knowledge in a domain through practice). A recurrent observation is that experts have strong metacognitive skills (Hatano, 1990). While learning and problem solving, they engage in regulatory behaviors such as knowing when to apply a procedure, planning, predicting the outcomes of an action, questioning the limitations of their knowledge, monitoring their progress, and efficiently apportioning cognitive and emotional resources. The capacity to regulate one own's learning and adapt accordingly further distinguishes routine experts from adaptive experts, with the latter being "characterized by their flexible, innovative, and creative competencies within the domain" (Hatano & Oura, 2003, p. 28).

Research has also shown that general problem-solving procedures ("weak methods"), such as trial-and-error or hill climbing, are slow and inefficient (Pellegrino et al., 2001), and that experts instead use deep knowledge of the domain ("strong methods") to solve problems. This deeper knowledge does not refer to isolated facts, but rather to knowledge encoded in a way that closely links it with its contexts of practice and conditions of use. When experts face new problems or scenarios, they can readily activate and retrieve the subset of their knowledge that is relevant to the task at hand (Simon, 1980). The implication for assessment here is that learning and skill mastery cannot readily be divorced from their specific contexts of practice, and as such, measuring deeper learning processes must occur in context rather than in highly generalised problem contexts.

Design innovations to measure learning and support teaching

In this section, we highlight five broad design principles that we argue should form the basis of "next-generation assessments" that can yield potentially valid evidence about where students are in the development of 21st-century competencies (in summative applications) and about what they need to do to progress in these skills (in formative applications). These principles reflect our increasing understanding of the processes involved in developing higher-order thinking and learning skills, thanks to extensive research in the learning sciences, as well as the types of experiences that elicit these competencies

Design Principles

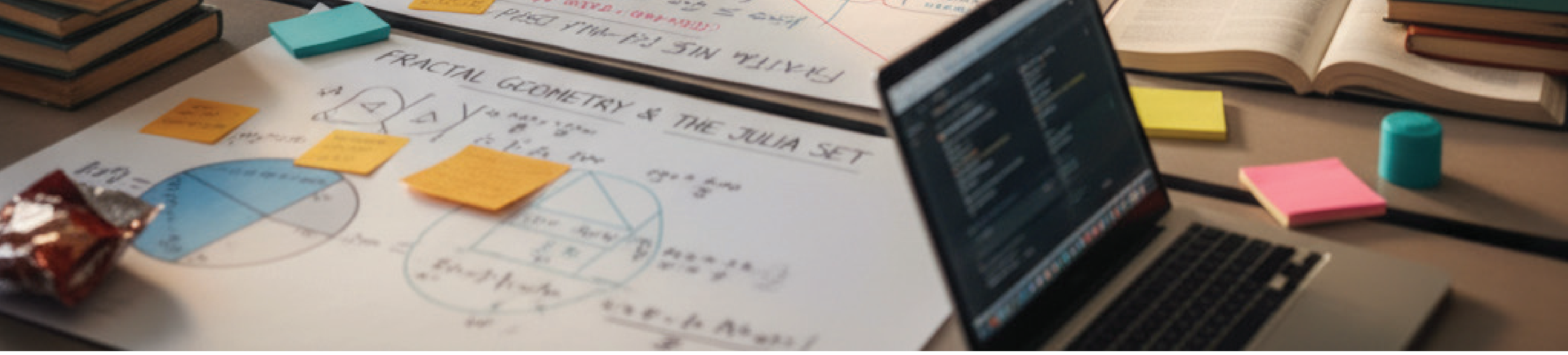
01 Include Extended Performance Tasks

02 Account for Knowledge in Task Design and Reporting of Performance

03 Provide Opportunities for Productive Failure

04 Provide Feedback and Instructional Support to Students During Tasks

05 Design Tasks That Are 'Low Floor, High Ceiling' or Otherwise Adaptive



01

Design Principle 1:

Include Extended Performance Tasks

Assessments that aim to measure how prepared students are for deeper learning—and to generate insights useful to teaching and learning processes—must engage students in active and authentic learning processes. From an assessment perspective, this means providing students with a purposeful challenge that replicates the key features of those educational experiences where deeper learning happens. Evaluating students' capacity to construct new knowledge in choice-rich environments means that students should be given sufficient time and affordances to demonstrate what they can do. We argue that extended units with multiple activities, sequenced as steps towards achieving a main learning goal, can provide a more authentic and motivating experience of assessment that provides valid data about students' competencies (i.e., evidence that is predictive of what students can do outside of the constrained and stressful context of a test).

These environments can facilitate a more open interpretation of students' goals and their exploration of the problem constraints, reward diverse solution strategies and outcomes, and provide feedback to learners. This also makes it possible to observe metacognitive processes that are crucial to learning in a non-obtrusive way, tracking how students plan and implement strategies, how they behave when they are stuck, and how they respond to feedback (Nunes et al., 2003). The application of a principled design process can lead to a productive use of these process data to augment the evidence that is derived from final solutions, therefore reducing the trade-off between reliability and authenticity (e.g., Piacentini, 2023; Sabatini et al., 2023).

02

Design Principle 2:

Account for Knowledge in Task Design and Reporting of Performance

Knowledge plays an important role in authentic task performance. Competencies like creativity, critical thinking or communication are rarely exercised within a vacuum. In an assessment context, students' ability to demonstrate these skills will always be observed within a given context and their knowledge about this context or situation will influence the type of strategies they use as well as what they are able to accomplish (Mislevy, 2018).

To make valid conclusions about students' skills, assessments must account for the knowledge they bring to each task.

What this means, in turn, is that it is important to explicitly identify the knowledge that students need in any assessment context to meaningfully engage with the test activities and to evaluate the extent to which differences in prior knowledge influence the evidence we can obtain on the target constructs. It is also important to assess complex skills across a variety of knowledge and application domains to make valid conclusions on students' mastery of these skills. Evaluating the extent to which students possess relevant knowledge when engaging with a complex performance task (for example, through a short battery of items) should become an integral part of the design and assessment process in next-generation assessments, as this can help to interpret their subsequent behaviors and choices during the assessment (Piacentini, 2023; Roll & Barhak-Mirkowitz, 2023).

Design Principle 3:

03

Provide Opportunities for Productive Failure

There is evidence that we can make robust claims about students' preparedness to learn new things by studying how they work on unfamiliar problems (Roll et al., 2011; Schwartz & Martin, 2004). For example, "invention activities" ask students to work on problems requiring concepts or procedures that they have not yet been taught, with the aim that students explore and understand the core properties of a construct before being taught expert solutions and strategies (Roll et al., 2012). Students often fail in their attempts to solve or generate canonical solutions to such problems, but experimental evidence shows that students who learn through invention activities are better at transferring their knowledge (i.e., solving other tasks requiring the same knowledge schemes in a different context) in comparison to students who are directly told how to solve the problem (Loibl et al., 2016; Kapur & Bielaczyc, 2012). In an assessment context, these types of activities can provide evidence about whether students can flexibly apply their knowledge schema to unfamiliar contexts.

In traditional tests, if students do not know the relevant procedure to follow there is little they can do to progress (Schwartz & Arena, 2013). Yet in the real world, we can access resources and ask more knowledgeable others for help when learning and problem solving. Assessments that challenge students to learn to solve new problems should incorporate resources for learning, because problem solving always requires some degree of knowledge. These should be carefully crafted so that they do not provide prescriptive solution steps, but rather provide opportunities to learn about core properties of the problem and encourage implementing a certain strategy that helps to make progress toward a solution.

Design Principle 4:

04

Provide Feedback and Instructional Support to Students During Tasks

To complement the design principle 3, instructional support in the form of advice, feedback, or prompts as students engage in activities can promote deep learning in beginners and enable the decisions they make in their learning to be observed (Azevedo & Aleven, 2013; van Joolingen et al., 2005). Targeted feedback and scaffolded interventions can also reduce the risk that beginners disengage from an assessment because they perceive it to be beyond their capacities: it can provide clarity over task instructions, reduce the degrees of freedom or number of acts required to make progress, signal critical features that a student may have missed upon first attempt, reproduce partial solutions, and elicit further articulation or reflection questions (Guzdial et al., 2001). All these are important functions that can serve to elicit more information about students' competency level and reengage students if they appear disengaged—which is especially important in the context of extended performance tasks with less discrete items and fewer clear "data points" to inform scores.

05

Design Principle 5:

Design Tasks That Are 'Low Floor, High Ceiling' or Otherwise Adaptive

Next-generation assessments of 21st-century competencies should enable all students to demonstrate their ability to learn and tackle problems by using the tools and resources available to them—regardless of their initial level of knowledge or skill. Adapting assessment challenges to different abilities not only improves the quality of the measures but also the authenticity of the assessment experience: in real life, people seldom take on challenges that are too easy or impossible to achieve, yet in traditional tests this happens quite frequently. One approach to catering to different student ability levels is to design so-called "low floor, high ceiling" tasks, meaning that they are largely accessible to all ability levels while providing scope to challenge top performers. One cluster of low floor, high ceiling problems asks students to produce an original artifact: for example, a story, a game, a design for a new product, an investigative report, a speech, etc. These artifacts generate a wide range of qualitatively distinct responses and even top performers are incentivized to produce a solution that is richer, more complete, and unique. This type of design can also be used in more standardized problem-solving tasks if students are informed about intermediate targets to achieve and that progress towards more sophisticated solutions will be rewarded.

Next-generation assessments should let every student demonstrate their ability to learn and solve problems using the tools available to them—just as they would in real life.

Adaptive designs can also address the complexity of measuring learning-in-action amongst heterogeneous student groups. A relatively simple way to integrate some level of adaptivity in assessment design is to structure achieving larger and more complex goals as a sequence of tasks that gradually increase in difficulty (similar to a "level-up" mechanism) and instruct students to complete as many as they can.

From principles to practice

These five design principles are intended as broad guidelines for designing next-generation assessments capable of measuring and supporting the development of 21st-century competencies. They are not intended to be prescriptive but rather illustrate the characteristics of innovative assessments that can provide valid information about students' real-life abilities. The two examples described in the full version of this chapter—the OECD's PISA LDW test and PILA platform—represent an initial engagement with the enterprise of innovating assessments (see Foster & Piacentini, 2025).

Designing assessments is a complex and challenging endeavor that must be guided by theory and research about cognition; yet just like any other design activity, scientific knowledge provides direction and constrains the set of possibilities, but it does not prescribe the exact nature of the design nor ingenuity in the final product. In the case of educational assessment, the design is influenced in important ways by variables such as its purpose (e.g., to assist learning, to measure individual attainment, or to evaluate a program), the context in which it will be used (e.g., classroom or large-scale), and practical constraints (e.g., resources and time) (Pellegrino, 2023). Our fundamental argument here is not to shift completely from one assessment paradigm (i.e., using only short discrete, static items) to another (i.e., using only extended, interactive performance tasks with resources and feedback), but rather encourage the development and use of a more diversified set of assessment experiences where the breadth and depth of tasks and associated measurement models are aligned with what the assessment intends to measure.

Concluding remarks and reflections for future work

This chapter has argued that the next-generation of assessments should focus on observing and interpreting how students solve complex problems and learn to do new things. Assessments often shape and influence the teaching and learning that takes place within education systems. If we want to shift policy into practice, we must be able to assess the competencies that students need for their future, meaning we must reproduce the features of these learning situations in assessments.

There is evidence that innovative assessments of educationally and socially significant competencies are both desirable and possible, and we have described some overarching characteristics of what these assessments should look like. What is also clear is that the vision described in this chapter cannot be undertaken without investment in multiple forms of capital: intellectual, financial and political (see Pellgrino et al., 2023, for a more in-depth discussion). The evidence suggests that cooperation and collaboration on a global scale, including through the design of open-source technology and model tasks, may well be the best and only way to achieve such advances—at least for the time being.

> ***If we want to shift policy into practice, we must be able to assess the competencies students need for their future.***

References

Azevedo, R., & Aleven, V. (Eds.). (2013). International handbook of metacognition and learning technologies. *Springer.* https://doi.org/10.1007/978-1-4419-5546-3

Binkley, M., Erstad, O., Herman, J., et al. (2011). Defining twenty-first century skills. In Griffin, P., McGaw, B., & Care, E. (Eds.), *Assessment and teaching of 21st-century skills.* Springer. https://doi.org/10.1007/978-94-007-2324-5_2

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2019). Implications for educational practice of the science of learning and development. *Applied Developmental Science, 24*(2), 97–140. https://doi.org/10.1080/10888691.2018.1537791

Darling-Hammond, L., Herman, J., & Pellegrino, J., et al. (2013). *Criteria for high-quality assessment.* Stanford Center for Opportunity Policy in Education.

DiCerbo, K., Shute, V., & Kim, Y. (2017). The future of assessment in technology-rich environments: Psychometric considerations. In Spector, J., Lockee, B., & Childress, M. (Eds.), *Learning, design, and technology: An international compendium of theory, research, practice, and policy.* Springer International Publishing. https://doi.org/10.1007/978-3-319-17727-4_66-1

Dumont, H., Istance, D., & Benavides, F. (Eds.). (2010). *The nature of learning: Using research to inspire practice.* OECD Publishing. https://doi.org/10.1787/9789264086487-en

Ericsson, K. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In Ericsson, K., Charness, N., Feltovich, P., & Hoffman, R. (Eds.), *The Cambridge handbook of expertise and expert performance.* Cambridge University Press. https://doi.org/10.1017/cbo9780511816796.038

Foster, N. (2023). 21st-century competencies: Challenges in education and assessment. In Foster, N., & Piacentini, M. (Eds.), I*nnovating assessments to measure and support complex skills.* OECD Publishing. https://doi.org/10.1787/e5f3e341-en

Foster, N., & Piacentini, M. (Eds.). (2023). *Innovating assessments to measure and support complex skills.* OECD Publishing. https://doi.org/10.1787/e5f3e341-en

Guzdial, M., Rick, J., & Kehoe, C. (2001). Beyond adoption to invention: Teacher-created collaborative activities in higher education. *Journal of the Learning Sciences, 10*(3), 265–279. https://doi.org/10.1207/s15327809jls1003_2

Hatano, G. (1990). The nature of everyday science: A brief introduction. *British Journal of Developmental Psychology, 8*(3), 245–250. https://doi.org/10.1111/j.2044-835x.1990.tb00839.x

Hatano, G., & Oura, Y. (2003). Commentary: Reconceptualizing school learning using insight from expertise research. *Educational Researcher, 32*(8), 26–29. https://doi.org/10.3102/0013189x032008026

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Loibl, K., Roll, I., & Rummel, N. (2016). Towards a theory of when and how problem solving followed by instruction supports learning. E*ducational Psychology Review, 29*(4), 693–715. https://doi.org/10.1007/s10648-016-9379-x

Mislevy, R. (2018). *Sociocognitive foundations of educational measurement.* Routledge.

Nunes, C., Nunes, M., & Davis, C. (2003). Assessing the inaccessible: Metacognition and attitudes. *Assessment in Education: Principles, Policy & Practice, 10*(3), 375–388. https://doi.org/10.1080/0969594032000148109

OECD (2018). *The future of education and skills 2030.* OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2018/06/the-future-of-education-and-skills_5424dd26/54ac7020-en.pdf

Pellegrino, J. (2023). Introduction: Arguments in favour of innovating assessments. In Foster, N., & Piacentini, M. (Eds.). *Innovating assessments to measure and support complex skills.* OECD Publishing. https://doi.org/10.1787/e5f3e341-en

Pellegrino, J., Chudkowsky, N., & Glaser, R. (Eds.). (2001). K*nowing what students know: The science and design of educational assessment.* National Academy Press. http://faculty.wiu.edu/JR-Olsen/wiu/common-core/precursor-documents/KnowingWhatStudentsKnow.pdf

References

Pellegrino, J., Foster, N., & Piacentini, M. (2023). Conclusions and implications. In Foster, N., & Piacentini, M. (Eds.). *Innovating assessments to measure and support complex skills*. OECD Publishing. https://doi.org/10.1787/e5f3e341-en

Piacentini, M. (2023). Defining the conceptual assessment framework for complex competencies. In Foster, N., & Piacentini, M. (Eds.), *Innovating assessments to measure and support complex skills*. OECD Publishing. https://doi.org/10.1787/e5f3e341-en

Piacentini, M., Foster, N., & Nunes, C. (2023). Next-generation assessments of 21st-century competencies: Insights from the learning sciences. In Foster, N., & Piacentini, M. (Eds.), *Innovating assessments to measure and support complex skills. OECD Publishing.* https://doi.org/10.1787/e5f3e341-en

Roll, I., & Barhak-Rabinowitz, M. (2023). Measuring self-regulated learning using feedback and resources. In Foster, N., & Piacentini, M. (Eds.), *Innovating assessments to measure and support complex skills*. OECD Publishing. https://doi.org/10.1787/e5f3e341-en

Roll, I., Aleven, V., McClaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*(2), 267–280. https://doi.org/10.1016/j.learninstruc.2010.07.004

Roll, I., Holmes, N. G., Day, J., & Bonn, D. (2012). Evaluating metacognitive scaffolding in guided invention activities. I*nstructional Science, 40*(4), 691–710. https://doi.org/10.1007/s11251-012-9208-7

Sabatini, J., Hu, X., Piacentini, M., & Foster, N. (2023). Designing innovative tasks and test environments. In Foster, N., & Piacentini, M. (Eds.), *Innovating assessments to measure and support complex skills. OECD Publishing.* https://doi.org/10.1787/e5f3e341-en

Schwartz, D., & Arena, D. (2013). *Measuring what matters most: Choice-based assessments for the digital age.* The MIT Press. https://doi.org/10.7551/mitpress/9430.001.0001

Schwartz, D., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics education. *Cognition and Instruction, 22*(2), 129–184. https://doi.org/10.1207/s1532690xci2202_1

Simon, H. (1980). Problem solving and education. In Tuma, D., & Reif, R. (Eds.), *Problem solving and education: Issues in teaching and research.* Erlbaum.

Shute, V., & Kim, Y. (2013). Formative and stealth assessment. In Spector, J., et al. (Eds.), *Handbook of Research on Educational Communications and Technology (4th Edition).* Lawrence Erlbaum.

Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2016). Rethinking assessment in a digital age: Opportunities, challenges and risks. *British Educational Research Journal, 42*(3), 454–476. https://doi.org/10.1002/berj.3215

Thornton, S. (2012). Issues and controversies associated with the use of new technologies. In Gormley-Heenan, C., & Lightfoot, S. (Eds.), T*eaching Politics and International Relations.* Palgrave Macmillan UK. https://doi.org/10.1057/9781137003560_8

van Joolingen, W., de Jong, T., Lazonder, A. W., Savelsbergh, E. R., & Manlove, S. (2005). Co-Lab: Research and development of an online learning environment for collaborative scientific discovery learning. *Computers in Human Behavior, 21*(4), 671–688. https://doi.org/10.1016/j.chb.2004.10.039

Zhai, X., Haudek, K. C., Shi, L., Nehm, R. H., & Urban-Lurain, M. (2020). From substitution to redefinition: A framework of machine learning-based science assessment. *Journal of Research in Science Teaching, 57*(9), 1430–1459. https://doi.org/10.1002/tea.21658

About the authors

Natalie Foster is an Analyst in the Programme for International Student Assessment (PISA) at the Organisation for Economic Cooperation and Development (OECD). Her work mainly focuses on the design and development of innovative assessments of 21st century competences included in each PISA cycle, working closely with measurement and test development experts, as well as various other PISA research and development projects. She is the lead author of the PISA 2022 Creative Thinking and PISA 2025 Learning in the Digital World assessment frameworks, co-editor of the publication Innovating Assessments to Measure and Support Complex Skills, and the lead author of the PISA 2022 Results (Volume III): Creative Minds, Creative Schools report. She has also worked in the OECD Centre for Educational Research and Innovation on the Smart Data and Digital Technologies in Education project, where she contributed to the OECD Digital Education Outlook 2023. Before joining PISA, she worked at the OECD Development Centre and European Commission.

Mario Piacentini is a Senior Analyst in the Programme for International Student Assessment (PISA) at the Organisation for Economic Cooperation and Development (OECD). An expert in measurement, Mario leads the work on the PISA innovative assessments and the broader PISA Research & Development Programme. He works with international experts to design assessments of 21st century competences. His projects aim to expand the metrics we use to define successful education systems. He is one of the authors of the Global Competence (PISA 2018) and Creative Thinking (PISA 2022) assessment frameworks, and he is currently leading the development of the PISA 2025 assessment of Learning in the Digital World and PISA 2029 assessment of Media and AI Literacy. He also coordinates the development of an open-source platform to support the use of technology-enhanced, formative assessments in the classroom. Before joining PISA, he worked for the Public Governance and the Statistics Directorates of the OECD, the University of Geneva, the World Bank and the Swiss Cooperation. He has authored several peer-reviewed articles and reports and was co-editor of the OECD publication on Innovating Assessments to Measure and Support Complex Skills. Mario holds a Ph.D. in economics from the University of Geneva.

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
November 2025

Citation
Foster, N., & Piacentini, M. (2025). Innovating assessment design to better measure and support learning. 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