Què funciona en educació?

Evidències per a la millora educativa

FUNDACIÓ JAUME BOFILL

ivàlua | Institut Català d'Avaluació de Polítiques Públiques

Integrating formative and summative assessment in a seamless system: how and why

Janet Looney, director of the European Institute of Education and Social Policy

Barcelona, 14 May 2018

Segueix el debat en directe via Twitter
#EduEvidències
@FundacioBofill @Ivalua_Cat

Key issues

- Definitions and concepts
- Barriers and tensions
- Strategies to improve integration

Definitions and concepts

The evaluation and assessment ecosystem

- External school evaluation
- School self-evaluation
- Teacher and school-leader appraisal
- National/international student assessments
- Classroom-based formative, summative and ipsative (learner self-referenced) assessment

Definitions and concepts

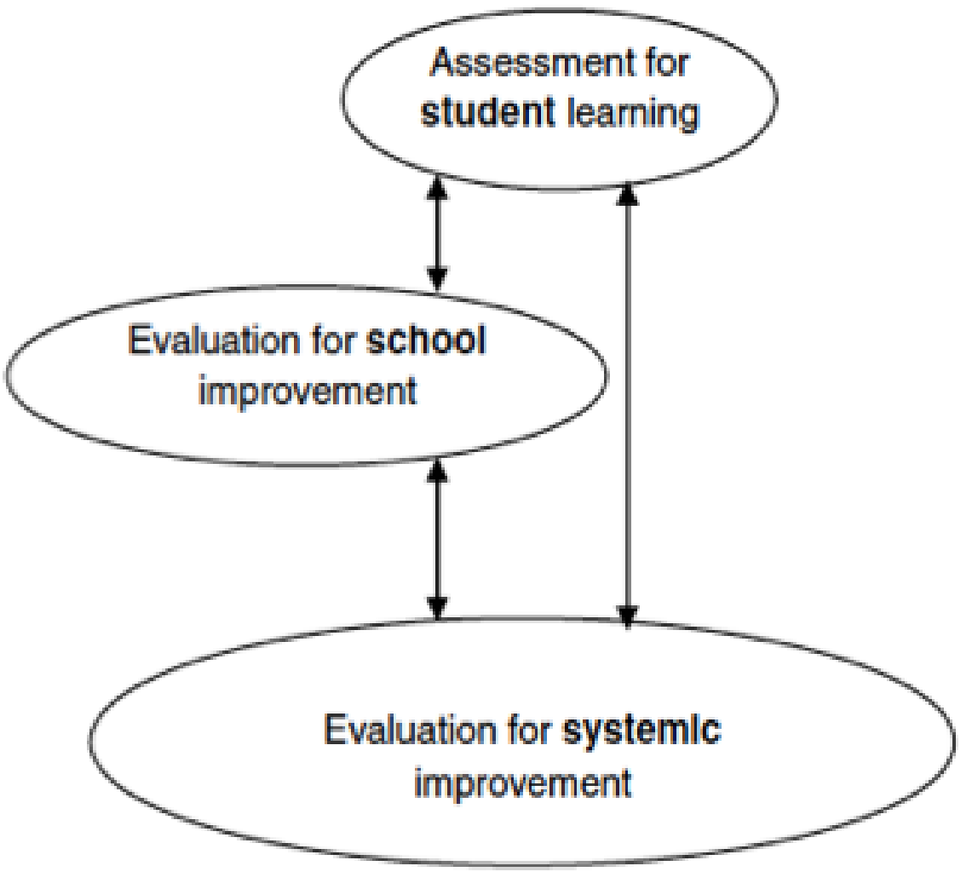
Types of assessment

- **norm-referenced**
- **criterion-referenced**
- **self-referenced (ipsative)**

Assessments of learning outcomes /competences are typically criterion-referenced, measuring attainment of standards, based on well-defined criteria.

- **Summative, e.g. at the end of a course or leading to certification**
- **Formative, e.g. to shape next steps in learning, or as feedback at the school or policy level**

Seamless integration means that assessment data may be used at every level of the system



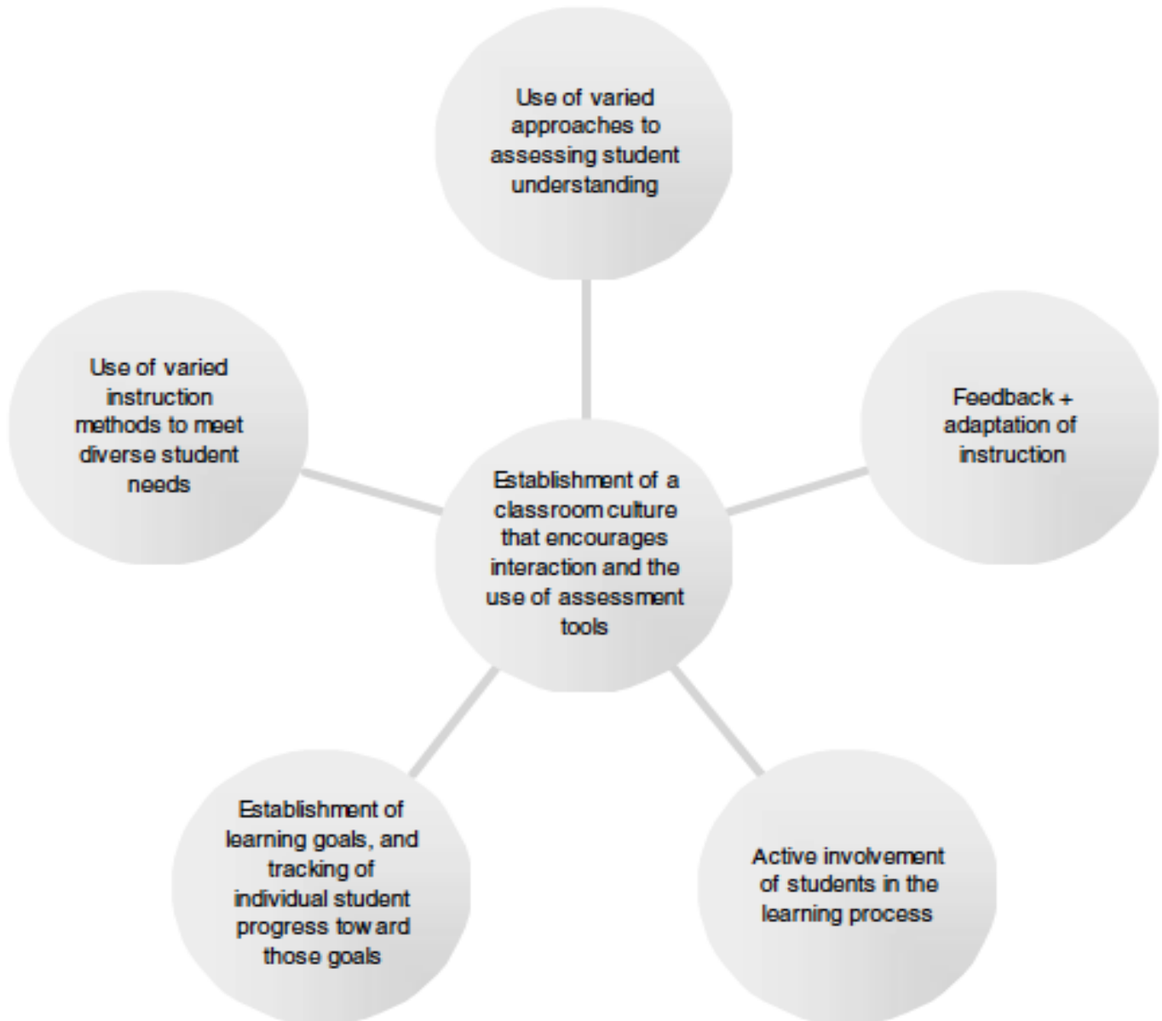
Source: OECD (2005)

Timing, detail (granularity) differ for each level

Formative assessment: A synthetic model

Formative assessment refers to the regular assessments of student progress to identify learning needs and shape next steps in teaching and learning

Figure 3.1. The six key elements of formative assessment



OECD, 2005

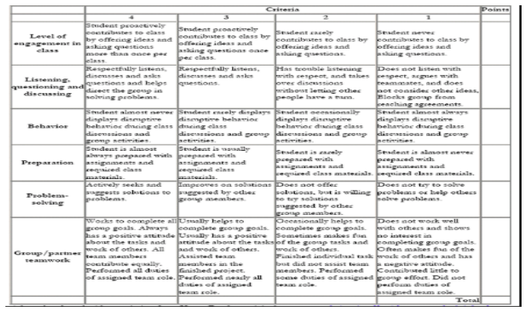
Integrating formative and summative assessment in a seamless system: how and why
Janet Looney

Large-scale summative assessments

Most common:
Large-scale digital assessments

Less common:

Portfolio assessments
& scoring rubrics



	Criteria				Points
	4	3	2	1	
Level of engagement in class	Student proactively contributes to class by offering ideas and asking questions more than once per class.	Student proactively contributes to class by offering ideas and asking questions once per class.	Student rarely contributes to class by offering ideas and asking questions.	Student never contributes to class by offering ideas and asking questions.	
Listening, questioning and discussing	Respectfully listens, discusses and asks questions and helps direct the group in solving problems.	Respectfully listens, discusses and asks questions.	Has trouble listening with respect, and takes over discussions without letting other people have a turn.	Does not listen with respect, agrees with teammates, and does not consider other ideas. Blocks group from reaching agreements.	
Behavior	Student almost never displays disruptive behavior during class discussions and group activities.	Student rarely displays disruptive behavior during class discussions and group activities.	Student occasionally displays disruptive behavior during class discussions and group activities.	Student almost always displays disruptive behavior during class discussions and group activities.	
Preparation	Student is almost always prepared with assignments and required class materials.	Student is usually prepared with assignments and required class materials.	Student is rarely prepared with assignments and required class materials.	Student is almost never prepared with assignments and required class materials.	
Problem-solving	Actively seeks and suggests solutions to problems.	Improves on solutions suggested by other group members.	Does not offer solutions, but is willing to try solutions suggested by other group members.	Does not try to solve problems or help others solve problems.	
Group/partner teamwork	Works to complete all group goals. Always has a positive attitude about the tasks and work of others. All team members contribute equally. Performed all duties of assigned team role.	Usually helps to complete group goals. Usually has a positive attitude about the tasks and work of others. Assisted team members in the finished project. Performed nearly all duties of assigned team role.	Occasionally helps to complete group goals. Sometimes makes fun of the group tasks and work of others. Finished individual task but did not assist team member. Performed some duties of assigned team role.	Does not work well with others and shows no interest in completing group goals. Often makes fun of the work of others and has a negative attitude. Contributed little to group effort. Did not perform duties of assigned team role.	
				Total	

Validity and reliability

- **Validity refers to degree to which assessments measure what they are intended to measure**
- **Reliability refers to the consistency, stability of results**

Results also need to be usable **- Usability refers to the ease with which results may be interpreted and used to make improvements**

Validity and reliability are fundamental for all types of assessment and at all levels

Validity and reliability in the context of authentic learning

Validity and reliability include authenticity and complexity of the task(s) in relation to the domain assessed, impact on student's learning (Gielen et al., 2003)

Achieving reliability is a challenge in authentic learning environments, but is possible

There are ongoing challenges in measurement of "soft skills" in large-scale assessments

Barriers and tensions

Barriers to seamless integration

Large-scale standardised assessments, which are designed to ensure that data are valid and reliable , cannot easily capture student performance on more complex tasks, such as problem solving, reasoning, or collaborative work (holistic vs. instrumental tension)

Large-scale assessments do not provide the detailed information needed to diagnose the specific sources of student difficulty (granularity)

Feedback needs to be timely and relevant to have an impact on student learning

In high-stakes contexts, assessments may focus teachers' attention on helping students to meet learning outcomes, but many teachers narrow instruction -- scores thus overstate students' performance

Effective classroom-based formative assessment requires capacity to orchestrate learning in new ways, to explore student thinking, to respond "on the fly", to support learners in developing their own assessment skills (holistic vs. instrumental tension).

Looney (2011)

Holistic vs. instrumental assessment

Holistic assessments – ability to solve complex problems, higher-order thinking

Behaviourist/instrumental approaches – measuring narrow learning outcomes

Tensions between holistic and instrumental approaches affect both summative and formative assessment

Different users and uses of assessment data

At the system level, aggregate data gathered periodically are adequate for decisions related to allocation of resources, to track student performance, equity, and so on.

In classrooms, teachers need more detailed and frequent information on student learning in order to respond to student needs.

Timing of feedback has an impact on learning

- **Long-cycle formative assessment: 4 weeks to a year or more**
- **Medium-cycle formative assessment: 3 days to 4 weeks**
- **Short-cycle FA: 5 seconds (on-the-fly) to 2 days**

(Shavelson *et al.,* 2008; Wiliam, 2004; 2006)

Wiliam (2004) found that medium- and short-cycle feedback had a much greater impact on student learning (over one year, double rate of students progress found in control classrooms.

Human rating of large-scale assessment

There is evidence that the validity and reliability and of assessment scores are quite high when human raters are well trained.

Participation in rating panels also provides teachers with valuable professional development experience.

At the same time, human rating systems are costly and time-consuming

Strategies to improve integration

Progress towards a seamless system

- Address teachers' incentives to "teach to the test" in high-stakes contexts
- Integrate multiple assessments of student learning over time
- Draw on advances in cognitive sciences to strengthen the quality of both formative and summative assessment
- Support research and development toward 3rd stage digital assessment
- Strengthen teachers' assessment roles

A design framework to support coherence

A design framework which embraces more than one desired purpose -- that is, formative and summative assessment -- at the outset, and which considers:

- **Cognitive demands, that is, the types of problem-solving, representation and procedural learning, as well as the content and situations to which they would be applied.**

- **Content boundaries involve the creation of ontologies, maps, or graphs, showing the key assessment content and the relationships among content or topics.**

- **Task characteristics boundaries operationalise what, how, and how much are presented to the examinee**

Parameters are systematically crafted with the help of content, learning and teaching experts.

(Baker, 2018; see also Mislevy et al., 2003)

Three stages in evolution of digital assessment

Bennett (2015) outlines the past, present and future of digital assessment

1st stage (basic):

- **Digitisation of traditional assessments; adaptive testing**

2nd stage (evolutionary change):

- **New formats (multi-media, short constructed responses, short essays, online interaction between test users,)**
- **Initial attempts to measure new constructs**
- **Automated item generation, online human scoring**

3rd stage (revolutionary change):

- Designed to serve both institutional and individual learning needs
- Designed from cognitive and theory-based domain models (evidence-centred design)
- Use complex simulations and other interactive performance tasks
- More integrated with instruction, and sample performance repeatedly over time

R&D in digital assessment of complex competences

O'Leary et al. (2018) describe 3rd stage R&D :

- **Virtual reality (VR) simulations, with scoring based on experienced rater's observations, R&D on valid, reliable automated scoring**
- **Their most significant characteristic is that decisions about design, content and format are informed by competency models and by general cognitive principles from learning sciences**

VR creates realistic contexts; tasks are presented in a progressive fashion and encourage learners to apply assessment criteria as part of the assessment process

Virtual Reality (VR)

Game-based assessments

May create more authentic contexts, support learner self-regulation, collaboration, immediate feedback, challenge and competition

(Hess and Gunter. 2013)

Games may provide a tailored learning experience (design, curriculum sequencing adapted to prior knowledge, learning goals) and problem-solving support (feedback, hints to scaffold learning)

BUT, research and development for games that bridge entertainment and pedagogical purposes are still in early stages

(Kickmeier-Rust and Albert, 2010)

See also Groff, 2018

Strengthen teachers' assessment roles: a holistic approach

- **Structured activities**

  Broaden teacher repertoire and ability to respond "on the fly" to diverse student needs

- **Structured discussions (dialogic assessment)**

  Deep questioning

  Ensure teacher capacity to identify learner misconceptions, engage with students in their reasoning processes

- **Feedback (task-oriented, timely)**
- **Test banks to support teachers' summative assessment, training to support reliability of assessments**

***What matters is how teachers handle responses. Teachers need to collect student ideas, summarize and challenge them* (Black, 2007, 2018).**

References

Baker, E. (2018), “Design for Assessment Change”, *European Journal of Education*, Vol 53, Vol 2, pp. 138-140

Black, P. (2007), “Full Marks for Feedback”, Journal of the Institute of Educational Assessor, Spring, pp. 18 – 21.

Black, P. (2018), “Helping Students to Become Capable Learners”, *European Journal of Education*, Vol 53, Vol 2, pp 144-159

Gielen, S., F. Dochy and S. Dierick (2003), “Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects”, pp. 37-54 In Segers M., Dochy F., & Cascallar E. (Eds.), Optimizing new modes of assessment: In search of qualities and standards, Kluwer Academic Publishers, Dordrecht.

Groff, J.S. (2018), The potentials of game-based environments for integrated, immersive learning data, *European Journal of Education*, Vol 53, Vol 2, pp. 188-201

Hess, T. and G. Gunter (2013), “Serious game-based and nongame-based online courses: Learning experiences and outcomes” British Journal of Educational Technology, Vol. 44, No 3, pp. 372 –385

Kickmeier-Rust, M.D. and D. Albert (2010), ““Micro -adaptivity: Protecting immersion in didactically adaptive digital educational games,” Journal of Computer Assisted Learning), Vol. 26, doi: 10.1111/j.1365-2729.2009.00332.x

References, cont.

Looney, J.W. (2011), “Integrating Formative and Summative Assessments: Progress toward a Seamless System?”, OECD Education Working Papers, No. 58, OECD, Paris. doi: 10.1787/5kghx3kbl734-en

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design*. ETS Research Report Series.

O'Leary, M., Scully D, Karakolidis, A. and V. Pitsia (2018) “ The state-of-the-art in digital technology-based assessment, *European Journal of Education*, Vol. 53, No. 2, pp. 160 – 175.

OECD (2005) Formative Assessment: Improving Learning in Secondary Classrooms, OECD, Paris.

Shavelson, R., et al. (2008), “On the Impact of Curriculum-embedded Formative Assessment on Learning: A Collaboration between Curriculum and Assessment Developers.”, Applied Measurement in Education, Vol. 21, pp. 295-314.

Wiliam, D. (2006), .“Formative Assessment: Getting the Focus Right.”, Educational Assessment, Vol. 11, pp. 283-289.

Wiliam, D., C. Lee, C. Harrison and P.J. Black (2004), .“Teachers Developing Assessment for Learning: Impact on Student Achievement.”, Assessment in Education, Vol. 11, pp. 49-65.

Thank you!

Free trial for articles from the European Journal of Education

bit.ly/EJEfreetrial