

---

## System-level assessments

### BRIEF 4

#### Quality of education

#### Teacher improvement

#### Inspection

#### Classroom observation

**System-level assessments can improve learning if they are designed to ensure validity, reliability, and equity, and if the data is used to inform systemic change.**

Summative assessments conducted at regional, national, and international levels can serve two major purposes: certifying students' academic achievements and monitoring and evaluating educational provision and quality at a systemic level. In developing summative assessment systems, the different purposes of summative assessment should be considered, as well as the key properties of content, validity, reliability, quality-assurance, and impact.

### **Issues and Discussion**

What is Assessment for System Monitoring? Most education systems collect information about students' learning through regional or national examinations. However, examinations are typically used more for certification and selection of individual students, than for monitoring the quality of the education system as a whole. Increasingly, countries monitor the quality of the education system through a separate programme of testing or surveys that involves samples of pupils at certain ages or grades. These assessments do not give scores or feedback to individual students, but rather provide aggregated results for measuring trends over time.

Examples of such assessments for systemic monitoring include national and regional tests in specific subjects (e.g. [Finnish education evaluation plan 2012-2015](#)); citizen-led household-based assessments (e.g. the [basic](#) learning assessments conducted in India, Pakistan, Kenya, Tanzania, Uganda, Mali, and Senegal); or international school-based tests such as the Trends in International Mathematics and Science Study ([TIMSS](#)), the Progress in International Reading Literacy Study ([PIRLS](#)), the Programme for International Student Assessment ([PISA](#)), the studies conducted by the Latin American Laboratory for Assessment of the Quality of Education ([LLECE](#)), the Analysis Programme of the CONFEMEN Education Systems ([PASEC](#)), the Southern and Eastern Africa Consortium for Monitoring Educational Quality ([SACMEQ](#)), and the Southeast Asia Primary Learning Metric ([SEA-PLM](#)). These various assessment instruments are markedly different from one another,

---

and while they are usually low-stakes for pupils and their schools, they are high-stakes and often high-cost for governments, politicians and policy makers. Hence, in-depth knowledge of the functions, methods and limitations of each type of assessment proves essential to education planners whose decisions rest on these test results.

**Designing Large-scale Summative Assessment for System Monitoring:** As planners decide whether to join an existing assessment regime or design their own, there are a number of key questions to keep in mind. These include: What is the purpose of the assessment and how will the results be used to inform practice? What curriculum area (e.g. maths, science, mother tongue) or what construct (literacy, numeracy) needs to be assessed? At what stage should assessment take place and for what purpose? How often should the assessment take place? What standardised instruments, administration and scoring procedures should be used? What costs need be covered and by whom? Should the assessment be for the total population or a representative sample? How will individual student achievements in different schools/regions be aggregated to the system level? What other background information (school resources, family characteristics) should be collected to analyse the final results? How will the results be analysed, shared, and made use of? While these questions cannot all be answered here, four fundamental considerations should be kept in mind for all processes of assessment design: content, validity, reliability, and quality assurance.

- *Content:* The content of an assessment is determined by which aspects of learning are most valued and what can be expected of students in terms of progression and achievement. Good understanding of educational objectives and clearly defined learning outcomes will help identify the types of tasks to be included in the assessment or portfolios of evidence.(6)(11)
- *Validity:* The overall validity refers to whether an assessment accurately measures what it is intended to measure. Not all knowledge can be broken down into tasks that fit the process and time constraints of common summative assessment methods. Ideally, assessment tasks should complement each other to measure knowledge in context, application, analysis, and capability.(2)
- *Reliability:* A reliable assessment measure produces consistent results across related items within the same test, across different instances of test administration, and across the scores assigned by different raters. The reliability of an overall assessment system lies in the methodology used in student sampling, the design of the assessment instruments, administration and scoring procedures, and methods of data aggregation and analysis. Reliability also implies that assessment must be consistent and comparable across candidates, with minimisation of bias and error from assessors.(11)
- *Quality Assurance:* In those countries where corruption and political nepotism remain part of the social context, transparency and quality assurance procedures become paramount.(3) Generally, the more weight given to the summative assessment, the more stringent the quality assurance system needs to be through such methods as inter-school moderation in scoring, double marking, machine-markable tests, and the use of special software to analyse the results across cohorts, schools and regions.(8)(16)

**Impacts of Summative Assessment:** At the national, regional, and school levels, data from summative assessments is one of the most important sources of information for analysing the performance of the education system, diagnosing problems, projecting trends over time, determining system-wide or targeted intervention, planning institutional capacity building, and teacher training, designing curricula, and budgeting and distributing resources.(10) In many countries, assessment data is also used to hold accountable and improve the performance of schools and teachers.(9) While there may be positive impacts in terms of teachers working harder and more effectively to prepare all their pupils, accountability pressure may also impact negatively on learning and teaching, such as by

---

restricting teaching methods and content, and by reducing teacher morale.(12)(14)

At the level of learners themselves, summative assessment can be constructive for individual pupils when assessment tasks embody the desired learning outcomes, and when feedback on the results is used for formative purpose (e.g. summative assessment + feedback = formative assessment).(15) When pupils are involved in assessment processes they develop a better understanding of learning goals and they are primed for higher cognitive engagement in progressing towards these outcomes.(4)(5) However, an education system that places great emphasis on summative assessment and selectivity produces students with strong extrinsic orientation towards grades and social status, but weak intrinsic motivation for longer term learning.(2) Moreover, frequent use of high-stakes tests and examinations may lead to exam cheating and pressures to expand learning time through the [private tutoring industry](#).

## Inclusiveness and Equity

Inequality and Social Injustice: Summative assessments interact in complex ways with social injustice, inequality, deprivation, and other forms of disadvantage.(1) Disadvantaged learners who have low self-esteem and confidence, or who lack motivation and commitment, can be further demotivated by the pressure of tests and examinations. Furthermore, misuse of the aggregate results by the media and politicians can do considerable damage by consolidating unfair and inaccurate stereotypes. Many disadvantaged students can benefit greatly from more personalised modes of summative assessment, such as project and portfolio work, which can foster high levels of engagement.(1)(11)

## Policy Examples

- Finnish [\[PDF\]](#)
- Uganda [\[PDF\]](#)
- Netherlands [\[PDF\]](#)
- SEAMEO Region [\[PDF\]](#)

## References and sources

1. Beets P. and van Louw, T. (2011) [Social justice implications of South African school assessment practices](#), *Africa Education Review*, 8(2): 302-317
2. Boud, D. (2000) [Sustainable Assessment: Rethinking Assessment for the Learning Society](#), *Studies in Continuing Education*, 22:2, 151-167
3. Bethell , G. and Zabulionis, A. (2012) [The evolution of high-stakes testing at the school–university interface in the former republics of the USSR](#), *Assessment in Education: Principles, Policy & Practice*, 19:1, 7-25
4. Carless, D. (2015) [Exploring learning-oriented assessment processes](#). *Higher Education*, 69(6), 963-976.
5. Carless, D. (2007) [Learning-oriented Assessment: Conceptual Bases and Practical](#)

- 
- [Implications](#). *Innovations in Education and Teaching International*, 44(1): 57-66
6. Harlen, W. (2007) [Assessment of Learning](#). London: SAGE
  7. Harlen, W. (2009) [Improving Assessment of Learning and for Learning](#), *Education 3-13: International Journal of Primary, Elementary and Early Years Education*, 37(3):247-257
  8. Harlen, W. (2012) [On the Relationship between Assessment for Formative and Summative Purposes](#), In Gardner, J. (Ed.) *Assessment and Learning*. London: SAGE
  9. Hutchinson, C. and Young, M. (2011) [Assessment for Learning in the Accountability Era: Empirical Evidence from Scotland](#), *Studies in Educational Evaluation* 37: 62–70
  10. Kellaghan, T., Greaney, V. and Murray T. (2009) [Using the Results of a National Assessment of Educational Achievement](#). Washington DC: The World Bank.
  11. Lau, A.(2015): [‘Formative good, summative bad?’](#) - A Review of the Dichotomy in Assessment Literature, *Journal of Further and Higher Education*.
  12. Lee, J. (2008) [Is Test-Driven External Accountability Effective? Synthesizing the Evidence from Cross-State Causal-Comparative and Correlational Studies](#), *Review of Educational Research*, 78(3): 608-644
  13. R4D. 2015. [Bringing Learning to Light: the Role of Citizen-led Assessments in Shifting the Education Agenda](#). Washington, DC: Results for Development Institute
  14. Stobart, G. and Eggen, T. (2012) [High-stakes Testing – Value, Fairness and Consequences](#), *Assessment in Education: Principles, Policy & Practice*, 19(1): 1-6
  15. Taras, M. (2010) [Assessment for Learning: Assessing the Theory and Evidence](#). *Procedia Social and Behavioral Sciences* (2): 3015–3022
  16. Zupanc, D., Urank, M. and Bren, M. (2009) [Variability analysis for effectiveness and improvement in classrooms and schools in upper secondary education in Slovenia: Assessment of/for Learning Analytic Tool, School Effectiveness and School Improvement](#). *An International Journal of Research, Policy and Practice*, 20(1): 89-122

[print](#)