

Workshop: Towards a New Generation of Standardized Student Assessments

Executive Summary

The *Instituto Nacional de Evaluación de la Educación* (INEE) and the World Bank organized a workshop with the objective of contributing to the design of the second generation of standardized student assessments in Mexico. Extensive discussions with international and domestic experts took place from June 23 to 24, 2014 covering experiences and lessons from different countries and viewpoints from the academic field and public policy practice. Some of the workshop's main messages are presented below:

1. Assessments can have a formative or a summative purpose.
2. Formative assessment is supported by a curriculum-based design of the test and of teacher training. These two elements can improve the implementation of a formative assessment by (i) avoiding teaching to the test, cheating and gaming and, (ii) incentivizing the use of results as an input for improvement.
3. Summative assessments can be used to improve accountability and transparency, allocate resources across the system, and evaluate the effects of public policies.
4. Countries need to strike a balance between the formative and summative purpose of the test. Many pitfalls arise when one test is expected to fulfill several purposes.
5. The optimal design of student standardized assessments is subject to the characteristics of the education evaluation system as a whole and the presence of an enabling environment.
6. To achieve continuity in student assessments – a very important element determining success – the purpose and use of the information generated from the tests should be widely disseminated and the authorities should design strategies to evaluate the impact of assessments.

Assessments can have a formative or a summative purpose. Formative assessments give quick and timely feedback regarding students' performance to enhance teacher practices. Teachers benefit from students' formative evaluation since it helps them ensure that students' results improve over time and correct mistakes throughout the school year, thus improving learning. Formative evaluation is usually the responsibility of each teacher and it is therefore important to ensure that teachers are adequately trained to develop formative tests and use their results. On the other hand, summative assessments provide information on the status quo of the education system and its performance, providing useful and objective feedback to a wide range of stakeholders including school principals, teachers, parents, etc. Summative evaluation is usually the responsibility of school directors or education authorities.

Formative assessment is supported by a curriculum-based design of the test and teacher training. These two elements can improve implementation of a formative assessment by (i) avoiding teaching to the test, cheating and gaming and, (ii) incentivizing the use of results as an input for improvement. To better align student

assessments with the education system, the national curriculum (when available) should be taken into account in the development of standardized tests and the results from these tests should also help revise the curriculum. There is also a need to train and further engage teachers in the design and use of formative assessments and the use of the information provided by both formative and summative assessments to ensure that student assessments help improve learning. Unintended negative consequences, such as teaching to the test, gaming, and cheating can be avoided if the adequate “empathic” accountability approach is followed and the necessary teacher training is provided. A comprehensive and sound training in this area appears to be lacking internationally and in Mexico.

Summative assessments can be used to improve accountability and transparency, allocate resources across the system, and evaluate the effects of public policies.

Accountability interventions are commonly perceived to be those that tie teachers’ or directors’ incentives (salaries, promotions, or school closure) to tests results. There are, however, other ways of promoting accountability without such “high stake” measures. For instance, providing timely and clear information on learning outcomes to different stakeholders for them to identify challenging areas can help promote accountability and improve student learning more effectively. In addition, more “empathic” accountability avoids the perverse incentives that are typically created under more “high stake” approaches: teaching to the test and cheating, among others.

Countries need to strike a balance between the formative and summative purposes of the test. Many pitfalls arise when one test is expected to fulfill several purposes.

It is a complex task to try to use one single test to achieve both formative and summative objectives. There are always tradeoffs in test design. For example, a complex and more comprehensive test (using open questions, written essays, etc.) can usually better and more accurately measure learning; however, they are hardly comparable from year to year and cannot be applied to a large number of students. Designing a student assessment system with more than one test can help balance the formative and summative goals in evaluation.

The optimal design of student standardized assessments is subject to the characteristics of the education evaluation system as a whole and the presence of an enabling environment.

Student assessments are more likely to reach their objectives if they are part of an integrated and comprehensive education evaluation system including supervisors, school directors, and teachers, among others. It is important to take into account that teachers’ practices cannot be assessed and improved with only the results of student assessments. For instance, teachers’ practices can be monitored via classroom observation tools, student surveys, teaching portfolios, etc. In addition, while decisions regarding goals, design, implementation, and use of student assessments can be strongly based on academic and public policy expertise, it is crucial to take into account the contextual factors that will ultimately drive success in the system. Authorities should work on building an evaluation culture where all stakeholders focus on student learning and ensure that the information provided by the evaluation system helps improve learning. In

this sense, it is desirable to pilot test student standardized assessments before scaling them up whenever possible.

To achieve continuity in student assessments – a very important element determining success – the purpose and use of the information generated from the tests should be widely disseminated and the authorities should design strategies to evaluate the impact of assessments. A continuous effort to assess the student evaluation system is important to adjust and improve student assessments and, for instance, ensure that the results are real and not inflated due to potential corruption of the tests. Constant adjustments should be made to cultivate a dynamic accountability of the evaluation system. In this regard, it is equally important to empirically assess the theory of change of the evaluation system to gain an understanding of the levers that can ultimately help drive success.