

## A 'wicked problem': using high stakes testing of student learning in development – Part 2

**Author :** Robert Cannon

**Date :** May 6, 2013

When making comparisons about the quality of education, many people seem convinced of the validity of the data from tests such as the 'Trends in International Mathematics and Science Study' ([TIMSS](#)) and the 'Programme for International Student Assessment' ([PISA](#)), a study of 15-year-old's performance on mathematics, science and reading.

However, the suitability of such tests for [international comparisons](#) is [disputed](#). For this reason, and the risk of them becoming 'high stakes' tests with potentially negative consequences for developing countries, I argued for caution in the use of such tests in an earlier [post for this blog](#).

Research ([here](#) and [here](#)) on the effects of the Australian NAPLAN tests shows that it is becoming a high stakes test and, some argue, that it is having negative impacts on curriculum, schools, teachers and children. What are the implications of similar testing programs in developing countries with much weaker capabilities in educational management? Can they use the data from TIMSS and PISA for planning and quality improvement?' If the US is not [considered](#) ready to engage with PISA, then surely caution is imperative in less developed countries.

Here, the idea of '[capability traps](#)' in developing countries is illuminating. Those working in development often import 'standard responses' to address local problems, in this case, international tests of student learning. This can lead to the adoption of the features of well-functioning organisations and systems that may conceal a continuing and deeper lack of capability in the developing country.

I concluded my [earlier post](#) by promising to explore these issues further.

### What can be done?

World Bank educationist, Pasi Sahlberg, [notes](#) a trend towards an increased focus on literacy and numeracy in school curricula, including in schools participating in international development programs. This trend illustrates an emphasis on structural knowledge and technical skills in many countries, evidence, he believes, of an imbalance at the expense of beliefs, values and morality – concepts that receive emphasis in developing country school curricula such as in Indonesia. Little wonder these countries perform poorly on international tests compared with others that give more emphasis to structural knowledge.

However, it seems inevitable that there will be continuing demand for data from international tests. It is less clear what can be done to ameliorate the potential risks for developing countries. Possible strategies include:

*Consider the relationship between national education goals and the information that international tests can*

*validly provide.*

Review with development partners such questions as: Do proposed tests adequately measure important national educational goals? How are the frequently stated commitments to equity in education reconciled with the major tests [actively excluding](#) students with disabilities from being measured in ‘desired target populations’? How, exactly, is it planned to use the data from tests?

*Being realistic about what tests can achieve*

Avoid unreasonable international comparisons, sardonically discussed [here](#) as ‘PISA-envy’.

Challenge the flawed thinking associated with an assumed connection between testing and learning improvement. There is no evidence in the massive [review](#) of educational research by John Hattie that international comparative testing is having the impact implied.

Review the experiences of developed nations with tests. Consider what Nichols and Berliner, leading US assessment specialists and critics of high stakes testing, have to say [here](#): “The scores we get from high stakes tests cannot be trusted – they are corrupted and distorted. In addition, these tests cannot adequately measure the important things we really want to measure. This leads to the obvious conclusion that a moratorium on their use is needed immediately (Chapter 7)”.

Recognise that the problems may not only be with the tests themselves, but with the use of the data from the tests and with the consequences of test outcomes for children, teachers and schools.

*Donor leadership*

When the assessment of student learning is being considered, donors can exert influence by demanding, from project contractors and consultants, information about how the effects of any assessment programs are to be monitored and managed.

*Professional development*

Strong professional understanding among development specialists of assessment and the issues around the impacts of assessment is essential. The assessment literacy of consultants, advisers, teachers and administrators is especially important in the context of developing country reform and quality improvement.

*Public scrutiny of assessment*

During the preparation of development projects, we need to analyse the potential impact of testing in each national context – and in regional contexts as well, such as Papua in Indonesia – before embarking on programs of reform. Acknowledge that national and international tests can become high stakes tests and that such testing can create serious consequences. Identify these consequences and manage them to minimise harm, as discussed in [The Paradoxes of High Stakes Testing](#).

When educational reform is being considered, we would do well to think about adapting the strategies of environmental impact assessments to assess the possible positive and negative impacts that a proposal might have.

Continuing, independent [monitoring](#) of assessment is also essential so that all known consequences – positive as well as negative – can be identified and addressed. We have much to learn about this approach from other professions, such as medicine.

*Consider alternatives designed for developing countries*

The UNESCO e-book, *Smaller, Quicker, Cheaper: Improving learning assessments for developing countries*,

[published](#) in 2011, is a guide to the ways that learning assessments could be undertaken in developing countries. It argues for culturally sensitive assessments calibrated to policy goals and costs, with priority for positive impacts.

Another potentially fruitful approach is to consider the ideas in the most recent edition in the series *Education in the Asia-Pacific region: Issues, concerns and prospects*, Volume 18, 2013, *Self-directed learning oriented assessments in the Asia-Pacific*. [This volume](#) has numerous articles that examine the ways in which assessment methods are being reformulated in the region.

### **Final thoughts**

As frustrating as it will be for development specialists, it is not clear that we yet have the appropriate knowledge about assessment to conduct high quality tests for comparisons and for quality improvement across diverse cultures.

And, I am sure we have less than adequate appreciation of the great complexity of the issues. The use of international testing in education is truly a 'wicked problem' described [here](#) (page 286) as "...a poorly formulated social problem where available information is confusing, where decision-makers hold conflicting values and where proposed solutions often turn out to be worse than the symptoms".

*This is the second in a two part series on high stakes testing. The first part can be found [here](#).*

*Robert Cannon is an associate of the [Development Policy Centre](#).*