Validity, Test Use, and Consequences: Pre-empting a Persistent Problem

By Madhabi Chatterji on March 17, 2014 12:05 AM | No comments

In 2000, London opened its Millennium pedestrian bridge to the public in a widely celebrated event. The momentous occasion drew large numbers of people, eager to view and experience the bridge first hand. The numbers far exceeded the original expectations of organizers, a good thing. But, the bridge began to sway dangerously as greater numbers of people stepped on it. The underlying design specifications--drawn up by architects and engineers--assumed the bridge would carry much lower or differently distributed loads. They had not anticipated the huge crowd that appeared that day. Fearing a collapse, the event came to a premature halt and the bridge closed temporarily. [See **The Telegraph**]

Most well-developed standardized assessments in education are also designed to meet certain specifications. When test use extends beyond the assessment's specifications, "validity" of the information can suffer to different degrees, or threaten to collapse-- like London's Millennium bridge. In an ideal scenario, a test is engineered to tap into particular abilities or traits in a well-defined student population, with some clear purpose for the tool in mind. Before the tool is released for use, it should be tested and "validated" to ensure it can support the types of inferences and actions from test scores intended by users, meaningfully and reliably. When such actions involve test scores of individuals directly, certain types of studies would be initiated. When test scores are aggregated and statistically processed with additional types of data to make inferences about the performance of teachers, leaders or school organizations, other types of validation studies must be undertaken.

In a recent commentary in Education Week [**EW Vol. 33(24)**], I discussed validity and its relationship to test use with three illustrative cases. In this blog, I will discuss **consequences** in relation to **validity** and **test use**.

Test use has consequences in practice and policy contexts. Appropriate test use enhances the validity of inferences we make from test-based information, with consequences that are typically predictable and positive. Conversely, inappropriate test use undermines validity to different degrees, sometimes with undesirable consequences for concerned entities--whether they are test-takers, teachers, school leaders, or policy-makers.

If students learned more in school as a result of educational assessments tied to challenging content area standards, it would be an intended consequence of a testing program that would likely be viewed as a good thing by most, as it would be consistent with the overall goals of education. This, in fact, is an articulated mission of the current Common Core State Standards reform movement. To optimize validity of inferences about student learning here, curriculum, instruction and assessments must be well-aligned with the targeted learning outcomes and student expectations. For positive consequences to be realized, the common core tests should **not** be driving school reforms. Rather, the new curriculum and instruction that matches, should **precede** the administration of well-designed student assessments.

Negative consequences of testing programs are also not uncommon, some occurring without intention. Suppose culturally different student groups (say, due to a language barrier) are unable to respond to particular assessment items on a test tied to their placement in a special educational or job preparation program. As a consequence of the language interference, their test scores would be lower and invalid to some degree. In the context of a larger program admission policy, an unintended negative consequence could also occur. Should the validity issue remain undetected and uncorrected, these students could be barred inadvertently from receiving subsequent program benefits that better-performing students (for whom the test scores were more valid) would automatically receive. Such a consequence, if widespread, may even be viewed as negative for society in that it would prevent more equitable distribution of opportunities for all, resulting in under-development of human potential.

Many observers have documented other negative consequences of high stakes sanctions and rewards tied to results of assessment programs in school evaluation and accountability contexts, such as "teaching to the test", unwarranted or premature school closings that fail to account for contextual factors like a neighborhood's poverty, or forced school management changes towards a more corporate approach that contradicts core values of public education.

We should remember that not every assessment program has adverse impacts. Further, timely detection of adverse impacts can be mitigated by concerted efforts of measurement researchers as well as practitioners and policy-makers working together. This, I hope, is an ideal that we could all strive towards.

What we need is more formal training, discussion and cross-learning among these key constituent groups to improve understandings of how common validity issues arise, with a view towards pre-empting unintended and adverse consequences in future. Enhancing validity and reducing adverse impacts in applied settings should be a shared responsibility of all assessment actors: test developers and researchers in testing programs, to be sure; but, educational policy-makers, teachers, practitioners, and public test users, as well.

More discussions on these and related assessment issues by measurement and policy scholars from around the world can be found here [See **Validity and Test Use**] and here [See **TCR Vol. 115 (9)**]. Policy briefs that translate the discussions to improve understandings of validity issues in practical settings, can be found here [**QAE Vol. 22 (1)**]. Thanks to Education Week for providing a platform for school officials and educators to engage in discussions with measurement and policy scholars on these issues.

Madhabi Chatterji
Teachers College, Columbia University

Categories: Validity   High-stakes testing   School accountability   International assessments   Test use

Tags: accountability   assessment   Common Core   PISA   SAT   test use   validity