

Sensitivity to instruction: the missing ingredient in large-scale assessment systems?

Prof. Dylan Wiliam

Institute of Education, University of London

UK

Educational policymakers all over the world rely on the results of large-scale accountability tests to inform policy. In doing so, they assume that the scores obtained by students indicate the quality of instruction the students have received—i.e., that the tests are sensitive to instruction. Using data from a variety of tests and examinations, this paper will establish that typical standardized tests are in fact not at all sensitive to instruction, for two reasons. The first is that the progress made by individual students is actually far less than the variability within an age cohort. The second reason is that the traditional processes of test construction decrease the sensitivity of a test to instruction, by systematically eliminating items that are sensitive to instruction. The paper concludes with two policy-relevant measures to address this. The first is that we should change the way we calculate reliability coefficients to prevent the systematic exclusion of items that are sensitive to instruction, and the second is a public information campaign to raise awareness of the issue of sensitivity to instruction, so that users of accountability test results understand the limitations of these tests as measures of the quality of education provided.