

Book review

Monitoring Educational Achievement, Postlethwaite, T., 2004. Paris: UNESCO Institute for Educational Planning, ISBN:9280312758 (139pp)

International comparative studies of educational achievement have increased recently in terms of numbers, resources devoted to them and the amount of attention paid to them by the media and policymakers. A volume that attempts to explain them, their strengths and weaknesses, to non-experts ought to be welcomed, especially when written by one of the key figures involved in their 50-year history.

The author sets out the following aims for the book:

1. To justify the importance of monitoring of achievement with examples.
2. To discuss and respond to criticisms of these international studies.
3. To point out the key technical aspects that should concern users.
4. To set out the implications of such studies for national policymakers at whom the book is principally aimed.

Following a brief historical introduction, the first chapter is headed 'Why do countries undertake national assessments or participate in international assessments?' The author claims that the two principal reasons are to identify the strengths and weaknesses in a (educational) system, and to track changes over time.

In fact, what is described by the author is how achievement tests are conceived of and applied—for example, deciding whether a country has sufficient technical resources to carry out a large-scale study. The reader will look in vain for a general discussion of how test results might allow anyone to come to conclusions about system strengths or changes over time. Instead, the reader will have to make do with the examples described in the next two chapters.

The second chapter looks at two large-scale national studies, in Vietnam and Kenya. The Vietnam study is briefly described and some results presented. The method used extensively now by the International Association for the Evaluation of Educational Achievement (IAE) and OECD for their studies of defining 'skill levels' in a hierarchical series of increasingly elaborated statements is presented in a completely unproblematic fashion. Thus the author claims 'it was possible to establish six levels of skills in each subject matter' (p. 37). We are given no indication that there is indeed a debate about the validity of such definitions. (Let me declare here that I am among those who have published critiques of these studies that are not referenced in this volume.) The chapter goes on to quote results showing a high correlation between provincial average teachers' scores and pupils' scores on reading and maths test. Without any hesitation we are told that 'pupils taught by teachers with low skills in mathematics and reading had a serious handicap', as if it were really possible to adduce such a causal inference from the existence of a high correlation, and, moreover, one measured at province rather than teacher level. Similar levels of (surprising) naivety are woven into the whole fabric of this book. In the same chapter, there are references to choosing items to define a minimum skill level for 'the ability to survive in Kenyan society' (p. 49) with the reflection that some apparently poor results should encourage the reading specialists to think about redefining the levels. The notion that there can be a simple relationship between test scores and social functioning is just taken for granted, as is the idea that it is straightforwardly possible to make clear inferences about real changes over time from changes in test scores. At the end of this chapter, the author does make some attempt to deal with the problem of causal inference in discussing the relationships between achievement and, for example, region by adjusting for social background. In doing so,

however, he omits any mention of the vast amount of research that demonstrates that a minimum requirement for sound causal inference about such relationships is longitudinal data that allow adjustment for achievement prior to attending a stage or phase of schooling. Of course, the major national and international studies described in this book are cross-sectional and do not have such data—a fact which means they are really most unsuitable for adducing causality. To admit this, however, would be to undermine the whole rationale for this volume! I am not suggesting that studies of this kind are valueless, merely that they are able to answer only a limited number of questions.

The next chapter looks at some of the studies made by the Southern and Eastern Consortium for Monitoring Educational Quality (SACMEQ), the Programme for International Student Assessment (PISA) and IEA. Much is made of using SACMEQ to study achievement changes over about a 5-year period. Again we are treated to statements such as ‘the question may well be asked as to why there was a decrease. In order to answer this it would be important to have measures of change in other variables’. We are then provided with a list of changes over the same time period in other variables such as classroom resource items, whereas what we really need to know is whether adjusting for any changes in such variables affects the achievement changes. Then there is the issue of the comparability of the different SACMEQ tests 5 years apart: the reader is told that the tests were ‘equated’ as if that was an unproblematic procedure that can be left to technical experts rather than a highly contested area for debate. The PISA study is examined and some results quoted. Again differences between countries are given strong causal interpretations with clear ‘advice’ for policymakers (p. 67) and a summary is given of results that assume clear causal relationships.

The next chapter is headed ‘Criticisms of assessment studies and responses to criticisms’. Eight ‘criticisms’ are presented, yet some of the most

cogent are ignored. For example, there has long been a debate about translation issues, yet this is not mentioned. Likewise, the method of choosing and discarding test items is by no means unproblematic, yet this is not mentioned. On the whole, this chapter presents a very superficial picture; for example, under the heading ‘costs of such studies’, only the actual monetary costs are mentioned with no account taken of hidden costs in terms of teacher time etc.

The next chapter deals with technical standards by which studies can be judged. This is perhaps the most useful chapter in the book, providing a useful set of questions that a reader can ask about how a study has been conducted.

The final chapter raises some implications for educational planners. Once again the author ignores the key distinction between statistical associations and causality. While he remarks that ‘it is not the rank order (of the mean results) that is important’, he strikingly ignores the fact that it is precisely these rank orderings that get reported in the media and are taken up by governments. This is a pity, since Postlethwaite himself has been extremely influential in the whole international testing movement. One of the problems, of course, is that the funding for these studies ultimately comes from governments who seem to be satisfied with the league tables that rank orderings produce.

Overall, it is difficult to recommend this book. Its treatment of the issues tends to be both superficial and misleading. It presents the issues as relatively unproblematic, and omits some of the more contentious ones. These studies are proliferating, their costs are increasing and governments take considerable notice of them. A serious and balanced evaluation is badly needed.

Harvey Goldstein
University of Bristol, Bristol, UK
E-mail address: h.goldstein@bristol.ac.uk