

Education for All: the globalization of learning targets

Harvey Goldstein*

University of London, UK

It is argued that the ambitious programme 'Education for All', launched by UNESCO, could be seriously undermined by its reliance upon the achievement of numerical 'targets'. Evidence from existing attempts by countries to impose educational targets reveals undesirable side-effects and distortions of educational systems. The paper explores some of the possible consequences of UNESCO's adherence to such targets.

Introduction

At its conference in Jomtien in 1990, and reaffirmed in Dakar in 2000, UNESCO adopted the declaration on 'Education for All' (EFA) that has subsequently become one of its major programmes, with implementation plans at least up to the year 2015. Full details can be found at www.unesco.org/efa. The basic aims, set out in six goals, encourage all countries, but especially those in the developing world, to implement policies resulting in certain basic educational 'standards'. These are linked directly to other concerns, including economic progress and health, via a stated belief in the necessity of basic education as a foundation for satisfactory progress in these other areas. A particular emphasis is on access to primary education with the eradication of gender differences and those based upon class and ethnic status. There is also a concern with adult literacy. An implementation programme is being developed, in conjunction with individual countries and regional and international organizations such as the World Bank and certain NGOs.

In order to focus its activities, EFA has set up targets for achieving its aims by certain dates and the present versions of these were agreed at the Dakar meeting. Thus, for example, it is intended that by 2015 'all children ... will have access to and complete free and compulsory primary education of good quality', and by 2015 there will be 'a 50% improvement in levels of adult literacy'. In this article I look closely at these targets, their epistemological status, and some of the consequences of pursuing them. My intention is not to provide a general critique of the programme's aims, which are widely accepted, but rather to explore ways in which the pursuit of particular kinds of targets could undermine these same aims. In particular I will focus upon the 'learning targets' such as those concerned with adult literacy

*Bedford Group, Institute of Education, University of London, 20 Bedford Way, London WC1H 0AL, UK. Email: h.goldstein@ioe.ac.uk

and children's achievements. My main thesis is concerned with the distorting effects that 'high stakes' target setting can lead to, by encouraging individuals to adapt their behaviour in order to maximize perceived rewards; viewed as a rational response to external pressures and for which there is now a body of research evidence derived from existing educational systems, notably in England and the USA. Other targets, for example concerned with enrolments, may also be subject to the same effects, but will generally have less direct effect upon the learning process itself, and I will not discuss these.

Measuring targets

The sixth Dakar goal refers to 'recognized and measurable learning outcomes' being 'achieved by all'. There is, however, no indication of the problematic nature of such measurements and no clearly set out description of what form the relevant assessments might take. The goal of obtaining a 50% improvement in adult literacy similarly lacks detail. Responsibility for setting up the necessary instruments and implementing them appears to have been assigned to UNESCO's statistics division based in Montreal. I shall first look at existing evidence on the measurement of adult literacy and then at the issue of targets for primary schoolchildren.

The most recent and best-studied literacy survey is the International Adult Literacy Survey (IALS) that represents the collaboration of a number of countries who agreed cooperatively to investigate adult literacy on an international basis. The main findings are published in a report (OECD, 1997) and there is also a technical report (Murray *et al.*, 1998). Nine countries initially took part, five EU member countries (France, Germany, Ireland, the Netherlands and Sweden) together with the USA, Canada, Poland and Switzerland.

A draft report of the results of the IALS in December 1995 revealed concerns about the comparability and reliability of the data, and the methodological and operational differences between the various countries. In particular, France withdrew from the reporting stage of the study and the European Commission instigated a study, including a re-analysis, of the EU dimension of IALS. The results from that investigation are reported elsewhere (Carey, 2000). The ostensible aim of IALS was to provide a comparison of levels of 'prose', 'document' and 'quantitative' literacy among the countries involved, using the same measuring instrument that would yield equivalent interpretations in the different cultures and different languages. Respondents, about 3000 in each country, were tested in their homes. Each participant responded to one booklet, which contained items of each literacy type, and there were seven different booklet versions which were rotated.

The re-analysis identified several problematic aspects of IALS, and by implication all such attempts at international comparisons (Blum *et al.*, 2001), including attempts to set common international targets. In these comparisons the dominant paradigm is that of a common measuring instrument that allows 'comparable' scores to be obtained from individuals in different educational systems. This operates through a process of translation and certain psychometric scaling techniques.

The first issue concerns the problems of translation and cultural specificity. As Blum *et al.* (2001) point out, there are some things that are culturally or educationally specific so that exact translations are impossible, and in many cases it is not possible to predict in advance which these items are. If a measuring instrument is restricted only to those items for which we might assume there are no locally specific differences, there is then a real question about whether such an instrument is measuring anything useful. To illustrate their point, Blum *et al.* (2001) give several examples. One of these concerns whether charts presented to respondents in a particular format that is more familiar in some countries than in others. Other examples are concerned with the ways in which particular languages embody linguistic structures that favour certain kinds of question wordings. Through a re-analysis of data they demonstrate how the actual IALS items produced biased responses in particular cases. One example that illustrates the subtlety of cultural bias is worth repeating. The task requires respondents to work out which are the comedies in a review covering four films. In two of these reviews, in both English and French, the term 'comedy' appears, which makes the question easy. In France, however, they find that many interviewees gave as their answer a third film, which from the description is obviously not a comedy. The only possible explanation is the presence in that film of the actor Michel Blanc, who is well known in France for his roles in many comedies but is little known abroad. Here, association predominated in the answering process to the detriment of careful reading of the reviews.

The second issue concerns the ways in which, for each of the three aspects of literacy, a single score was derived from test item responses using psychometric techniques based upon the assumption of a single underlying 'dimension'. As is typically done in international comparative studies, little attempt is made to explore the existence of more than one underlying dimension, despite evidence (see Goldstein & Browne, 2004, for an example) that such dimensions do exist. Blum *et al.* (2001) point out that the psychometric scaling used tends to remove items from the final test instrument that do not 'fit' the unidimensionality assumption. They point out that the initial *balance* of items representing different dimensions therefore determines crucially what the final test actually measures and that this distorts interpretations. Nevertheless, having just a single dimension does allow a simple rank ordering of countries and the subsequent publication of international 'league tables'. Thus, the political requirement is satisfied by the application of a particular technical (psychometric) model.

The third issue, which is especially pertinent for EFA, is in how literacy levels are defined. IALS, for example, uses a complicated series of five 'levels', from basic to advanced. Blum *et al.* (2001) demonstrate that there are alternative, and arguably equally valid, formulations that lead to very different views about the 'problem' of low literacy levels. For example, Blum *et al.* investigate the use of a measure of literacy level based upon the 'best' response given by a respondent rather than the 'average' response. Distributions of literacy level, using this measure, are completely different from the IALS distributions. Using the IALS measure, 65% of French interviewees with non-imputed scores have a prose literacy level of 1 or 2 (the lowest levels) while for a measure based on the best response, the proportion falls to 5%.

For Great Britain, the proportions are respectively 48% at level 1 or 2 using the IALS measure, and 3% at the same level using the alternative measure. Thus the EFA goal of improving literacy by 50% is strictly meaningless unless a particular definition is adopted and justified. Indeed, using the Blum *et al.* (2001) alternative measure with the above values it would also be somewhat absurd. The failure of EFA to recognize and articulate this issue suggests that the stated aim has more in common with a political slogan than with a scientifically based aspiration.

Targets, high stakes and teaching to the test

Experience within existing educational systems shows that an emphasis on numerical learning targets can be dysfunctional. My argument is that similar considerations will apply internationally—in particular, that any rise in test scores should not be confused with a rise in learning achievement as opposed to test-taking performance.

In England, a system of defined achievement targets for children at different ages was set up by the 1997 New Labour government. Evidence has accumulated that while there has been a general rise in actual test score levels in those aspects of the curriculum tested and in public examination results, highly dysfunctional consequences have emerged. One of these is the tendency to demotivate pupils and increase their test anxiety, especially among low achievers (Harlen & Crick, 2002). Teachers also feel that their professionalism is undermined, including their capacity for creative innovation. Radnor (2002) summarizes the research evidence. One result of the controversy over this issue is that the devolved administrations of Scotland, Wales and Northern Ireland have all decided to abolish the publication of league tables.

Interestingly enough, the obsession with numerical targets eventually created serious problems for government ministers. Thus Stephen Byers, then School Standards minister, in January 1998 claimed that 57% (in 1996 for English for 11-year-olds) ‘will not do—that is why we have set a target of 80% by 2002’ (Byers, 1998). It turned out to be 75%, although by then Byers had been moved elsewhere. The Secretary of State at the time, David Blunkett, in fact staked his job on achieving such targets, but sensibly transferred departments before having to confess to failure. His successor as Secretary of State, Estelle Morris, was a party to the original claim and she resigned in 2002, at least partly because the target was not reached. At the end of 2002 the government admitted that many of its earlier targets had been missed (DfES, 2002). These were almost all learning targets and included targets for examinations at 16 and 19 years of age as well as those for younger students.

The second source of evidence on this latter issue comes from the USA where there is gathering evidence that ‘high stakes’ testing systems that reward schools or teachers on the basis of their pupils’ test scores can certainly improve test scores but may have no effect upon learning which is assessed independently of those tests. In the state of Texas, under former Governor George Bush, a very high-profile testing programme was instituted in 1990 for grades 3–10 in Texas schools. The results are used to rank schools in league tables and certain funds are allocated on the basis of

the test results. Over the 1990s very large gains in student test scores were observed, and certain ethnic minority differences were reduced. Dubbed the ‘Texas miracle’, these results have been used as a justification for such testing programmes involving rewards given to schools for performance on the tests.

The most important manifestation of this trend in the USA is the ‘No Child Left Behind’ Federal Education Act of 2001 (www.nochildleftbehind.gov/) which mandates testing of all school pupils in grades 3 to 8 and publication of results in league table form. In one important respect it goes further than legislation in England by giving parents the right to transfer a child from a low-scoring school to a higher scoring one. There has been some strong opposition to this Act from teacher unions among others (see for example, <http://www.nea.org>).

Researchers from the RAND corporation have compared the results of the intensive testing programme in Texas with results obtained from a national testing programme, the National Assessment of Educational Progress (NAEP) that is carried out over the whole of the USA (the report is downloadable from <http://epaa.asu.edu/epaa/v8n49/>). What they found was that for mathematics and reading, compared to the rest of the USA the comparative gain in test scores over time of the Texas students on the national test was much less than that implied by the Texas test scores, and in some cases no different at all from changes found in the USA as a whole. Moreover, the ethnic results from NAEP showed that, if anything, in Texas the differences were increasing rather than decreasing. The researchers conclude that the concentration on preparation for the Texas state tests may be hindering an all-round development of mathematics and reading skills, especially for minority students.

In both England and Texas, we see evidence that when learning outcomes are made the focus of targets, those who are affected will change their behaviour so as to maximize their ‘results’, even where this is dysfunctional in educational terms. At the international level it would not be surprising if we witnessed similar kinds of behaviour where the curriculum and educational infrastructures were manipulated to maximize performance on the international performance measures, whatever the deleterious side effects that this might produce.

Failing to meet the targets

While UNESCO has published lists of countries that it expects may miss the principal targets, little has been said about the consequences of so doing. If targets are to be meaningful then some kind of sanction has to be in place for those who fail to meet them. In England and in Texas these are clearly spelt out in terms of withholding resources or even closing down institutions.

In a strategy document (UNESCO, 2002) it is stated that ‘governments would demonstrate their commitment to education through efforts to transform their education systems, in response to which external partners would provide financial and technical support in a transparent, predictable and flexible manner’ (p. 38). One of these partners is the World Bank, which refers to the ‘aid-worthiness’ of countries (p. 36) and also to the target requirements in order for countries to ‘receive

significant increases in external financing and technical support' (p. 37). It seems therefore, that the targets are to be taken seriously and that the stakes are high in terms of aid and other support. This is just the situation therefore where one might expect the most 'at-risk', that is, poorest countries to be tempted into 'gaming' in order to maximize perceived external rewards.

The linking of aid to achievement of targets has already begun. In April 2002 the World Bank agreed a 'Fast Track Initiative' (FTI) for 'high-risk' countries (EFA Global Monitoring Report, 2002; World Bank, 2002). Under this, selected countries receive aid in return for achieving certain 'policy reforms'. These are associated with each country's 'macroeconomic, structural and social policies and programmes to promote growth and reduce poverty' (EFA, 2002), which are developed along with international agencies such as the World Bank and International Monetary Fund. It seems likely that the actual achievement of the targets will be linked closely to grants or loans and that the 'reform' of systems along particular lines will become a prerequisite for many countries.

The issue of providing technical support is also of some concern, since EFA does not specify how this will be done, or even what it means. Since, however, EFA involves both curriculum change and intensive assessment, it seems likely that international bodies will be invited to provide such support. Thus, curriculum development bodies, testing and examination bodies and textbook publishers are all likely players and these almost inevitably will be those with the most international experience. Thus, globally active organizations such as Educational Testing Service (Princeton, USA), University of Cambridge Local Examinations Syndicate (UCLES, Cambridge, England), CITO (National Institute of Educational Measurement, Arnhem, Netherlands), NFER-Nelson (London, England), and the Australian Council for Educational Research (ACER, Melbourne, Australia) may be expected to become a part of the EFA programme. Given the high level of technical expertise required to develop curriculum materials, and especially to construct suitable measuring instruments, we may also expect the direction taken by EFA to reflect, in part, the global interests of such corporations. Some of them are also closely involved with the Organization for Economic Cooperation and Development (OECD), sponsors of IALS, and increasingly becoming involved in international performance comparisons such as Programme for International Student Assessment (PISA) (OECD, 2001; download from <http://www.pisa.oecd.org>).

The operating rules in the General Agreement on Trade in Services (GATS) which are due to be finally ratified in 2005 (see <http://gats-info.eu.int/>) will make it much easier than before for these global corporations to establish their presence in local situations where there is little in the way of established expertise. They will be able to argue, for example, that they possess the psychometric methodology that is required to implement testing regimes. To challenge this, countries may be able to appeal, but such appeals will undoubtedly be judged by 'experts' who may well have links with, and who anyway can be expected to share the same assumptions and expertise as, the corporations themselves.

Conclusions

Given the deficiencies identified in the EFA targets, two important questions arise. The first is why the programme has advanced so far without serious attempts to deal with these problems. The second is whether it is possible to pursue the general aims of EFA without resorting to 'targets'.

The obsession of EFA with achieving specific learning targets seems to reflect a similar set of concerns within certain national education systems, as has been indicated with examples from England and the USA. Within such systems the imposition of targets for institutions or school authorities can be viewed as an effective means of centralized control (Radnor, 2002), even within the rhetoric of diversity and local decision-making. At the international level, even if unintended, the eventual outcome of pursuing EFA targets may well be an increasing control of individual systems by institutions such as the World Bank or aid agencies, supported by global testing corporations. The current designation of certain developing countries as unlikely to achieve the existing targets may not only lead to demoralization in those countries, it may also allow the imposition from outside of systemic reforms under the heading of 'remedies' to put those countries 'on track'. I have only been able to outline some of the forms this might take and further systematic analysis and understanding of the processes at work would be useful.

If we return to the general aims of EFA, and if we abandon learning target setting because of all the problems I have outlined, what then? If we accept the broad aims of EFA, to raise adult literacy levels and to raise quality and participation in primary schooling, then the really important issues are not those concerned with devising specific targets but those to do with the optimum ways in which these aims can be achieved. This implies that we need to find those alternative forms of delivery, curriculum design, pedagogy, financial incentives, etc. that work best *within each country*. Each educational system can develop different criteria for assessing quality, enrolment, etc. and instead of monitoring progress towards an essentially artificial set of targets EFA could concentrate the resources that it is able to mobilize towards obtaining the necessary understandings of the dynamics of each system. This would then allow constructive policies to be implemented. The emphasis would be on the local context and culture, within which those with local knowledge can construct their own aims rather than rely upon common yardsticks implemented from a global perspective.

Such a change would seem to run counter to the current orientation within UNESCO, which appears to derive from official philosophies of target setting and centrally determined 'benchmarks' which have prevailed within certain parts of the Anglophone world since at least the mid-1980s. Nevertheless, from the perspective of those countries identified as likely to fail to meet current targets, a locally contextualized perspective would seem to offer more potential for improvement. It is, after all, just those countries that are in most need of help.

Notes on contributor

Harvey Goldstein is professor of statistical methods at the Institute of Education, University of London. His principal scholarly interests are in the methodology of multilevel modelling and educational assessment.

References

- Blum, A., Goldstein, H. & Guérin-Pace, F. (2001) International adult literacy survey (IALS): an analysis of international comparisons of adult literacy, *Assessment in Education*, 8(2), 225–246.
- Byers, S. (1998) Press release, DfEE no. 005/98. Issued 7 January.
- Carey, S. (2000) *Measuring adult literacy—the International Adult Literacy Survey in the European context* (London, Office for National Statistics).
- DfES (2002) *Autumn performance report* (London, Department for Education and Skills) (www.dfes.gov.uk).
- EFA Global Monitoring Report (2002) *Is the world on track?* (Paris, UNESCO) (www.unesco.org/education/efa/monitoring/monitoring_2002.shtml).
- Goldstein, H. & Browne, W. (2004) Multilevel factor analysis models for continuous and discrete data, in: A. Olivares (Ed.) *Advanced psychometrics. A festschrift to Roderick P. McDonald* (Mahwah, NJ, Lawrence Erlbaum).
- Harlen, W. & Crick, R. D. (2002) *A systematic review of the impact of summative assessment and tests on students' motivation for learning* (London, EPPI-centre, Institute of Education).
- Murray, T. S., Kirsch, I. S. & Jenkins, L. B. (1998) *Adult literacy in OECD countries* (Washington DC, National Center for Education Statistics).
- OECD (1997) *Literacy skills for the knowledge society* (Paris, OECD).
- OECD (2001) *Knowledge and skills for life: first results from Programme for International Student Assessment* (Paris, OECD).
- Radnor, H. (2002) World class education in England: a possibility? *Research Intelligence*, 81, 12–21.
- UNESCO (2002) *Education for All: an international strategy to put the Dakar framework for action on Education for All into operation* (Paris, UNESCO).
- World Bank (2002) *Financing of first group of countries on Education Fast Track* (www1.worldbank.org/education/adultoutreach/index.efa.asp).