

Thermometer, pill, placebo or virus?

The ongoing debate on when to use standardised assessments and how

Martin Gustafsson¹ and Qetelo Moloji²

5 November 2011

Unfinished work-in-progress paper

Abstract

The chief concern of this paper is when and how to use standardised assessments in a schooling system to improve the quality of learning and teaching. The arguments made in the literature on this topic are intense and positions often seem irreconcilable. Positions are clearly informed by ideology but, less obviously, also by context, specifically regional and historical factors including recent experiences with standardised tests. A key contribution of this paper is a mapping of positions taken by analysts with respect to standardised testing against the context in which the analyst argues. It is expected that such a mapping can assist in the policy debates on the degree and type of standardised assessment to have within a specific context. The limited evidence available for or against the effectiveness of different forms of standardised assessment in different contexts receives attention. The academic literature in English from developed countries is considered, but also the grey literature and literature (some of it non-English) from developing countries, including African and Latin American ones.

1 Introduction

Manuals on how to conduct standardised assessments, policy guidelines on how to design rewards and incentives based on the assessments and analyses using test score data now feature prominently on the desks and computers of education planners in developing countries. These texts are to a fairly large degree produced by economists. Amongst education academics, positions on standardised assessments vary widely. Articles expressing deep scepticism or plain opposition to standardised assessments and what they entail are common in education journals. Similar standpoints are found in the policy documents of teacher unions, including the influential world federation of teacher unions, Education International. On those rare occasions when differing views on standardised assessments are expressed within the same conference, the debate easily becomes acrimonious and ideological. However, the education planners and economists, on the one hand, and the more pedagogically oriented educationists and teacher unionists, on the other, tend to hold separate conferences and rely on stock characterisations of the other side. Thus open conflict is avoided. This is of course itself a stock characterisation of the situation, yet one with enough truth to offer a point of departure.

This paper aims to examine the range of texts referred to above in order to make sense of apparent contradictions. This is something that appears to be seldom done. Whilst we do not claim to be impartial participants in the debate, an attempt is made to represent different positions fairly. The underlying assumption is that a careful assessment of the various positions can contribute towards a less fragmented and more informed debate.

¹ Researcher, Department of Economics, Stellenbosch University. Also advisor to the Department of Basic Education.

² Director, National Assessments, Department of Basic Education, South Africa.

Section 2 simply defines standardised assessments and traces their recent history. Section 3 looks critically at how ideology manifests itself in positions on standardised assessments. As will be seen, ideology, to a greater or lesser degree informed by evidence, plays an important role in debates on standardised assessments and education policy generally. Section 4 focuses on specific elements of standardised assessments and how policy choices might be influenced by, firstly, ideology and, secondly, circumstance. This section is important insofar as it finds that ideology is far more important with respect to certain policy issues than others and that often different ideological positions can lead to the same decision. Put differently, the planners and teacher unionists may share more common ground than they initially thought.

The paper focuses largely on developing countries and how these countries are influenced, sometimes in peculiar ways, by experiences in developed countries. Section 5 applies the framework sketched in section 4 to one developing country, South Africa, which finds itself at a critical point with respect to assessments, partly in order to demonstrate the framework. Section 6 concludes.

2 A short history of standardised assessments

Standardised assessments, of the both the international and national variety, originated in developed countries in the 1960s. The National Assessment of Educational Progress (NAEP) of the United States, established in 1969, offers the longest time series of any country's performance in standard tests covering key subjects, mathematics and reading in the case of NAEP. For each of three ages of pupils studied, every fourth year around 9,000 pupils are sampled. This translates into a sample of around 0.2%. This is sufficient to provide breakdowns according to a few key categories, including race. NAEP nowadays encompasses elements that go beyond the original scope, such as separate assessments aimed at providing reliable measures of pupil performance at the level of the state³.

The global TIMSS⁴ and PISA⁵ testing programmes, covering 64 and 74 countries respectively in their most recent runs, can also trace their roots back to the 1960s, specifically the First International Mathematics Study (FIMS) of 1964, which covered 11 countries. The global programmes are geared towards providing reliable comparisons of pupil performance across time and countries in key subjects, but they also generate important contextual data through background questionnaires completed by pupils, teachers and school principals. The following graph, which illustrates the proportion of the world's population in countries that have at some time participated in TIMSS or PISA, shows how the late 1990s saw a ballooning of developing country participation. Of the 60% of the world outside TIMSS or PISA currently, around two-thirds is accounted for by India and China⁶.

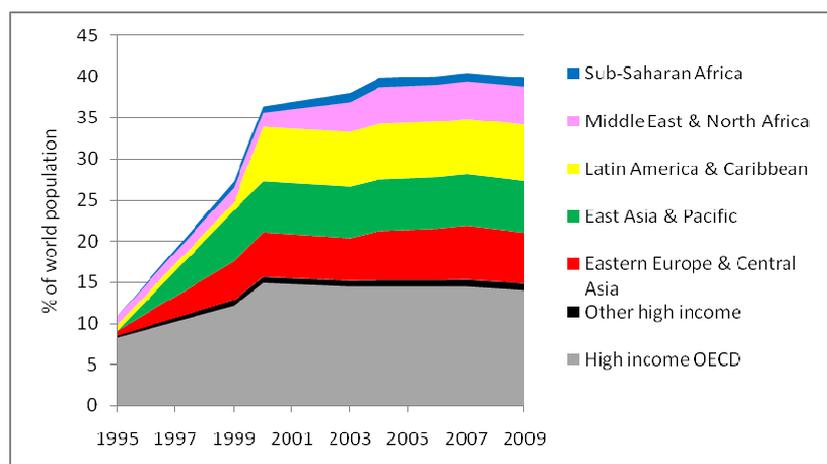
³ See <http://nces.ed.gov/nationsreportcard/>.

⁴ Trends in International Mathematics and Science Study.

⁵ Programme for International Student Assessment.

⁶ However, the city of Shanghai in China did participate in PISA 2009.

Figure 1: World coverage of TIMSS and PISA



Source: 2006 Revision Population Database of the UN Population Division. The country categorisations are those of the World Bank's World Development Report.

Data from the international programmes can of course serve the purpose of a national assessment programme of offering a time series of pupil performance. For many developing countries, these data are in fact the only source for such a time series. However, increasingly developing countries have established their own national programmes of the NAEP variety that are designed with the national curriculum in mind. The details of this trend are not well documented, but the 2007 Education for All (EFA) *Global Monitoring Report* tentatively lists 35 developing countries (UNESCO, 2006: 49) and a 2008 manual by the World Bank provides a few details on eight developing countries (Greaney and Kellaghan, 2008). Notably, India and China do not have national assessment systems. Occupying a layer between the global and national programmes are regional assessment programmes such as SACMEQ⁷ (14 African countries) and SERCE⁸ (16 Latin American countries), whose importance lies largely in the fact that they are able to differentiate sufficiently between levels of pupil performance typically found in developing countries (TIMSS and PISA, given their history, examine a range of performance that is more relevant for developed countries).

The spread of standardised assessments in developing countries is in part the result of the shift in emphasis within UN organs such as UNESCO from educational access understood as just enrolment to access understood as meaningful schooling that results in the acquisition of expected competencies. The milestone Jomtien World Declaration on Education for All of 1990 emphasised that 'It is ... necessary to define acceptable levels of learning acquisition for educational programmes and to improve and apply systems of assessing learning achievement' (UNESCO, 1990). This call has generally been interpreted as a call to establish new standardised assessment systems focussing on how well whole countries and possibly individual schools perform against agreed upon standards. It has not generally been interpreted as a call to strengthen or expand more traditional examination systems focussing on grading individual pupils. Why this should be so is not immediately obvious. Explanations include the fact that examinations as traditionally conducted do not provide system-level measures of educational quality that are comparable across time (Greaney and Kelleghan, 2008: 14). However, some of the arguments against using public examinations as tools to improve meaningful educational access, such as the fact that their 'high stakes' nature limit their ability to measure true educational progress, could also be made with respect to some national assessment systems. In short, the very strong emphasis placed in the literature on the use of standardised assessments in understanding and improving schooling systems has

⁷ Southern and Eastern Africa Consortium for Monitoring Educational Quality.

⁸ Second Regional Comparative Study (from Spanish).

perhaps been at the expense of a sufficient focus on how traditional examination system can be used and adapted for the same purpose⁹.

Much of the controversy around standardised assessments relates to the way systems covering all schools (as opposed to just a sample) are used to influence school behaviour. In the United States, the No Child Left Behind (NCLB) programme, introduced in 2001, requires states to conduct assessments in all schools and take certain actions in the case of poorly performing schools. NCLB operates rather separately from NAEP. A part of the reason for the strong emphasis in the literature on three Latin American countries, namely Brazil, Chile and Uruguay, is that these countries, but in particular the first two, now have standardised assessments with universal coverage linked to ambitious accountability reforms. This receives more attention in section 4.

3 Ideological undercurrents

In this section we sketch the ideologies of what we can call two opposing sides in education (later we shall see how easy it is to argue that this dichotomy is an over-simplification). We see how ideology finds expression in positions on education policy and, more specifically, standardised assessments. We assess the degree to which available evidence clashes with widely held beliefs. Finally, in keeping with our own ideology that identifying common ground in the debates is necessary, useful and possible, we attempt to do this.

Here we understand ideology as two things. Firstly, and broadly, it is a vision of the kind of social organisation that we should strive for. Secondly, it is a view of how education can contribute towards the desired social changes. We assume that at both levels ideology should be cognisant of what has been possible in the past, even if it is inevitably also informed belief in possibilities that have never been realised.

Before illustrative texts are discussed, we begin with our synopsis of the two sides. We label our first side the ‘teacher side’, because the positions are expressed to a large degree by teachers, teacher unionists and those who train teachers (of course we are not saying that all these people hold the position we are about to sketch). This side is on the political left and sees present inequalities very much as a problem of institutional structure. A powerful economic minority, which teachers are not a part of, maintains and expands its power through institutional means such as government policies which facilitate a type of capitalism that lends itself to large income disparities, national and global exploitation of people and unsustainable uses of the environment. Injustices persist even in the presence of democratic institutions, because those institutions are easily manipulated by the elite and, crucially for educationists, because people are deceived into accepting the status quo, as a result of their own political ignorance, or as the radical education academic Philip Wexler (1982: 57) put it, ‘the class-imposition of a false consciousness which contributes to social stability’. Democracy on its own cannot change things for the better, and must at the very least be combined with a critical level of political and economic understanding amongst the electorate. This, it is assumed amongst the more moderate, is enough to wrench control of institutions from the economic elite and to introduce policies that will sufficiently reign in capitalism in the public interest.

The young have an especially large interest in fundamental change in society and through their schooling are able to develop the interests and understandings required for this change and that the older generations lack. Teachers see themselves as critical agents bringing this change about. What makes this not just a project but a struggle is that the economic elite realises the threat to the social order posed by teachers and responds to attempting to manipulate the education system, through mechanisms of control developed in the private sector and through global collaboration between national elites, or globalisation, a term often

⁹ Bishop (1997) offers an important analysis of the impact of traditional examination systems.

used to encapsulate the forces against which teachers must struggle (Delandshere and Petrosky, 2004). Education International, the world federation of teacher unions encompassing, according to EI's website, 30 million teachers, is a powerful education policy force. Its recent education policy position paper describes the struggle as follows:

The social values of education require public authorities to protect the education sector from the neo-liberal agenda of privatization and commercialisation. This negative agenda includes marketisation and trade in education and intellectual property, the casualisation of employment in the education sector, the application of private-sector management models on education institutions, the privatisation of provision, and the intrusion of for-profit motives or business interest in the governance of education institutions. (Education International, 2011: 2)

We shall call the second side the 'public management side', again based on who tend to be the proponents of this position. As mentioned earlier, economists are prominent here. It is worth underlining that notwithstanding the strong polemics, many of the basic ideological tenets on the public management side are similar to those of the teacher side. Inequality is the pre-eminent social ill and education, more than any other activity, holds the key to overcoming inequality. However, the public management side sees social and economic inequality largely as a product of unequal levels of education. With better education, in particular in the form of improved writing, reading, critical thinking and mathematical skills, individuals will be able to capture more income within the capitalist economy. Economic inequality is seen to a large degree as a mirror image of educational inequality. In the public management view, again teachers play a central role in changing society, though the emphasis lies to a greater extent on realising higher levels of education and not at all on the role of the next generation as a political force bringing about change. The side characterised here does not always ignore the political factors that preoccupy the teacher side. Martin Carnoy, an education economist and one of the public management side's more left-aligned proponents, argues that globalisation can harm education, but that this is not necessarily true. What happens in the classroom is to a large degree impervious to larger economic changes, he argues, implying that to some extent fears amongst educationists are unfounded (Carnoy and Rhoten, 2002: 2). In fact, the public management side would argue that what is needed in the classroom is change, but good change, towards better levels of educational achievement, in particular for socio-economically disadvantaged pupils. This would be brought about to a large degree through management reforms and specifically incentives (not necessarily monetary) and sanctions aimed at ensuring that teachers do their best. If for the teacher side the enemy is an external political force, for the public management side the enemy is teachers not doing what they are paid to do, something that is widespread because teachers are often not required to account for what they accomplish in the classroom.

Do the two rather stylised representations appearing above contradict each other fundamentally? Or are we merely dealing with differences with respect to points of emphasis? One might be excused for thinking the latter. Contradictions with respect to the desired political economy should not surface as the public management side is relatively mute in this area. Would the teacher side really have a problem identifying the under-performance of poor pupils as a serious concern? Would the public management side have a problem with the view that we need more critically minded youths who are able to re-invent outdated institutions that stand in the way of a more just society? As we shall see, the divisions run deeper than these questions suggest.

Each side accuses the other of being a victim of ideology, unable to face up to important facts. The public management side stands accused of being naive, at best, or duplicitous in its role as a promoter of controls in education aimed at stifling critical thinking and undermining teachers in their role as agents of social change. Delandshere and Petrosky (2004: 2), in discussing the increasing use of standards in US schools, say the following:

A critical analysis of the reforms seems imperative, as they represent ideological stances on teaching and learning that have major consequences for the work of teachers... we raise here critical questions about their political, philosophical, and intellectual integrity and their intended and unintended consequences for teaching and for teachers' autonomy

The ideology of teacher side, on the other hand, is said to lie behind an exceptional lack of attention to facts and evidence in education policymaking. George Psacharopoulos (1996: 343), an economist attached to the World Bank who has been instrumental in developing economics of education as a field of analysis, has argued: 'In the field of education, perhaps more than in any other sector of the economy, politics are substituted for analysis.' Fernando Veloso (2009: xxv), one of twenty authors (19 of whom are economists) in a recent and influential book on Brazilian education reform, presents a similar complaint:

The debate around the distribution and use of public resources in education ought to be as prominent and meticulous as the debate around new health interventions. ... Prejudice adds an unnecessary layer of difficulty to the existing technical difficulties. ... Rhetoric, devoid of proof, ends up disqualifying empirical evidence wherever this runs contrary to pre-conceived notions of how the world ought to work. Curiously, the very people who reject this empirical evidence get their children vaccinated. (Own translation)

If one reads between the lines of Psacharopoulos, Veloso and others like them, the accusation one finds is essentially that there are powerful elements within the education authorities who oppose the measurement of pupil performance in order to deflect attention away from inefficiencies in the classroom, but also the education administration. The deification of teacher autonomy, the argument runs, is used as a means of keeping accountability out of the teacher profession.

Specific arguments for and against standardised assessments flow from the broader positions described above. Four broad objections to standardised assessments, or the way they are carried out, can be identified on the teacher side. Firstly, at a rather philosophical level there is discomfort with the message of certainty and predictability implied by the standards and strong emphasis on easily measurable knowledge, in particular mathematics, associated with standardised assessments. This detracts from more complex thinking and criticism in the classroom. This is not an accident, but deliberate:

Workers with critical thinking and problem-solving skills might not easily adapt to the economic requirements of business. In the face of uncertainty, contradictions, and the unknown, reformers have rallied around predefined standards, strict policies, and strategies, so that we have the illusion of knowing where we are, where we are going, and what and how we should teach. (Delandshere and Petrosky, 2004: 2)

Similar warnings are issued by Education International (2011: 4):

EI believes that a widespread abuse of the notion of quality to justify standardised forms of testing is harmful to the education system as a whole, as it attempts to reduce the teaching and learning process to quantifiable indicators. It is the standardization and one-dimensional approach to testing and evaluation of the teaching and learning processes to which EI objects strongly. ... Public authorities should guard against the potential misuse of the Programme for International Student Assessment (PISA) in the administration and planning of education systems.

Secondly, standardised assessments, as part of a larger package of management and accountability approaches, are said to be plain impractical in a school setting, apart from being ideologically suspect. This is the argument of the South African education academic, Yael Shalem (2003), who refers to inappropriate 'bureaucratic/managerial' methods employed by the education authorities to change schools through essentially more time spent by teachers on planning and reporting, according to a multitude of formats. Specifically with

respect to national standardised assessments, these might stimulate preparatory testing initiated by schools themselves, which then leads to more time administering tests and processing results and less time teaching. Thirdly, standardised assessments can serve as an easy placebo, creating the impression that something is being done to improve the quality of education, even though measurement on its own is clearly not the same as intervention. In particular, the argument is made, both in a rich country context (Delandshere and Petrosky, 2004: 6) and a developing country context (Shalem, 2003: 33) that education administrations use assessments to 'pass the buck' to teachers and thereby absolve themselves of their responsibility to lobby for better budgets for schools and provide support. Fourthly, there is something unpleasant and degrading for the professional teacher in having the thermometer of a standardised assessment being stuck into the classroom from time to time. Medical doctors do not experience similar assessments of the impact of their professional work on their patients, at least not at the level of the surgery, so why should teachers have to endure this? Shalem (2003: 30) invokes 'the great tradition of moral and ethical arguments in support of teacher professional autonomy' as a necessary counterweight to calls for more measurement.

The public management side, in contrast, welcomes more measurement, in particular in developing countries, where there tends to be little formal measurement and less educated communities are less able to keep teachers accountable in more informal ways. In fact, not having standardised assessments of some kind is considered a major flaw in a schooling system. As Bruns, Filmer and Patrinos (2009: 20), three World Bank economists specialising in education argue, 'It is hard to argue against greater access to accurate information'. How can there be no thermometer?

Luis Benveniste (2002: 106), another economist with a history at the World Bank, views the benefits of standardised assessments in a political light in discussing experiences in Uruguay (discussed further below).

In Uruguay, the introduction of assessment is embedded within a state model that advocates governmental intervention in the production and distribution of social welfare. Renato Operti asserts that the Uruguayan education reform efforts of the 1990s have been a "vindication of the Welfare State."

Standardised assessments, with their ability to 'call the bluff' when a state is offering poor quality schooling, are revolutionary in the sense that they can de-legitimise corrupt and ineffectual states that serve the interests of an elite. Bruns *et al* (2011: 35), using the economic notion of information asymmetries, argue that accountability systems such as standardised assessments, by empowering people with information on what the state actually offers them, can counteract the capture of public services by the rich. If information asymmetries are resolved, and people know what public services they should receive and what they actually receive, they have a much firmer basis for taking action to advance their rights.

Are the arguments of the two sides compatible with the evidence that we have on education and standardised assessments? Are there obvious failings against reality tests?

One critique that can be levelled against the teacher side is that sometimes, but not always, the ideological foe is presented more as a caricature than a real phenomenon. Such views are unlikely to be helpful in a struggle for a more just world, however one defines the latter. Resnik (2006: 175), in the *Comparative Education Review*, depicts the rise over the last half century of a deceptive and somewhat sinister project of world domination on the part of international organisations such as UNESCO and their education economists. These barely challenged forces are described as the new 'world education managers'. Undoubtedly some officials of the international organisations and a few economists would aspire to this level of power, but in reality, as Carnoy and Rhoten (2002) argue, education politics is complex, often fragmented and claims of the existence of global hegemonies tend to be exaggerations. To

some extent, the antipathy towards economics and, more broadly, quantification (which would encompass people such as education statisticians) is a rejection of the unknown, rather than a matter of fundamental ideological differences. It cannot be denied that sometimes economics and statistics are abused in a way that mystifies realities and gets policymakers to make the wrong decisions. There is a rich literature within economics about this problem, much of it revolving around Deirdre McCloskey¹⁰. Applying the criteria from this literature to economic analyses of education issues provides a more nuanced picture of when economics is helpful in education policymaking and when it is not.

Proponents of standardised assessments are themselves often concerned about the impact of globalisation, corporatism and managerialism in education. Greaney and Kellaghan, authors of prominent standardised assessment manuals of both the World Bank and UNESCO's International Institute for Educational Planning (IIEP), have also written about the risks of these assessments and acknowledge that these must be carefully weighed against the benefits of having the assessments (Kellaghan and Greaney 2001a, 2001b; Greaney and Kellaghan, 2008). Pragmatism is also evident in a paper by the Brazilian education academic Alda Castro (2008), who sees as the appropriate response to the spread of accountability systems in schools not attempting to stop the trend, but rather ensuring that these systems are designed to accommodate democratic participation and advance democratic governance generally. It is important to note that Education International (2011: 4) is not necessarily opposed to standardised assessments, but is opposed to their inappropriate usage:

When one form of evaluation designed for a particular purpose is used to serve a different purpose, the consequences can be unforeseen and damaging.

In South Africa, a recurring theme in the discourse between government and the country's powerful teacher trade unions has been whether support to teachers should precede standardised measurement of pupil performance and whether doing it the other way round is defensible. Surely you first ensure that teachers have the skills and tools they need before you begin measuring whether they can do their job. This has been the argument of the unions. Shalem (2003) pursues the same line of reasoning. But is this not a case of a false dilemma? Can support and measurement not be introduced simultaneously, with the guarantee that sanctions linked to the measurement will only be introduced after a period of sustained support? The problem with not introducing standardised measurement as soon as possible is that one can then not tell where to direct the required support.

A further problem with the discourse of what we have called the teacher side in South Africa, and in other developing countries, is an over-reliance on literature from developed countries with very different economic and social contexts. Shalem (2003), for instance, refers only to literature from the United States, the United Kingdom and South Africa (and those South African sources would tend to follow the same pattern in their sources). Below we shall explore some of the dangers of ignoring the literature from other developing countries.

Turning to the public management side, this side is often criticised for its over-reliance on models that put economic growth at the centre of national development and specify improved educational attainment (or quality) as the key determinant of growth. Many educationists balk at what they see as an overly narrow interpretation of education's role in society. In fact, there is a wealth of literature, including economic literature¹¹, that would support the educationists here. Moreover, promoting accountability reforms in schooling systems in no way requires the presence of an economic growth model. The economist Amartya Sen (1999: 5) argues as follows:

¹⁰ See for instance Ziliak and McCloskey (2008).

¹¹ Stiglitz, Sen and Fitoussi (2009) offer a recent (though pre-financial crisis), comprehensive and widely readable example of this.

...the opportunity to receive basic education or health care... are among the *constituent components* of development. Their relevance for development does not have to be freshly established through their indirect contribution to the growth of GNP or to the promotion of industrialization. As it happens, these freedoms and rights are also very effective in contributing to economic progress...

If more people can enjoy the pleasure of reading a novel this in itself is a sign of progress, regardless of whether this contributes to a country's gross domestic product or not.

A further critique that can be levelled at the public management side is that it does not always abide by its own rather stringent standards when it comes to data reliability and evidence. Despite a strong tendency to promote the national assessment systems of Chile and Brazil as examples of what developing countries should be pursuing, clear evidence that it is these reforms that lie behind improvements seen in, for instance, PISA, has not emerged. In the case of Chile, the economists have blundered a bit. Contreras, Flores and Lobato (2003) found, using data from the Chilean national assessments system and a linked incentive scheme, that the payment of incentives to teachers in schools showing exceptional results relative to the socio-economic status of the school community did improve subsequent pupil performance. However, as Mizala and Romaguera (2005) point out a couple of years later, this analysis violated fundamental rules around how to establish causality. Basically, if a school was on an upward performance trajectory, and received the reward for doing so, it might look as if receiving the award stimulated the subsequent improvements, but the critical causal factors may be the ones that lay behind the improvement trend occurring before the receipt of the award. Economists call complex two-way relationships between cause and effect endogeneity. In their own analysis that included better controls for endogeneity and data from a longer time period, Mizala and Romaguera (2005) found no causal effect between the rewards and better pupil performance. In the case of Brazil, as for Chile, arguments for these countries' major accountability reforms have had to rely on correlation rather than rigorous proof of causality. After the reforms were introduced major improvements in pupil performance occurred, therefore there is a possibility that the reforms caused the improvements, but we cannot be sure. Bruns (2010) emphasises that in the case of Brazil, which saw the second-largest improvements in mathematics in PISA between 2000 and 2009 (after Peru), reforms consisted of a simultaneous combination of pro-poor funding reforms and accountability and measurement reforms. It was this combination that brought about the improvements, she argues.

More rigorously derived proof of the effect of standardised assessments on pupil performance exists, in the case of the United States as a whole and specific regions within developing countries. The states within the US provide a quasi-experimental situation which Hanushek and Raymond (2005) take advantage of. They find that the timing of the introduction of pre-NCLB accountability reforms at the state level are sufficiently correlated to state-level performance improvements, as measured by NAEP, to establish a positive causal effect. Bruns *et al* (2011) take stock of the existing evidence in developing countries and find three cases where despite careful data collection and analysis, no clear causal link could be established between the standardised assessment (plus accompanying interventions and possible rewards) and improved performance (the cases are selected schools within Madagascar and the Indian states of Andhra Pradesh and Uttar Pradesh¹²). However, they also identify two important cases where the proof is irrefutable. In the Pakistan province of Punjab, the data point clearly to learning improvements in communities where adults were provided better information on how well children fare in school. Apart from improving information dissemination, there was no further intervention. In a group of schools in Liberia, better information to parents, based on standardised assessments, linked to a teacher capacity

¹² Bruns, Filmer and Patrinos (2011: 37, 58).

building programme, were found to lead to better levels of pupil learning (this case is discussed below)¹³.

Previously, it was pointed out that our two sides share some important common ground: both view social inequities as an ill to be combated through the use of education. To conclude this section, we point to a few areas where we are struck by an absence of explicit contradictions and believe that false or loose assumptions generate unnecessary polemic. We are not proposing that polemics are always redundant, nor are we denying that many education decisions are inherently charged by ideology. However, we do propose that through a more careful consideration of our assumptions we can conduct a more rigorous discourse, with a clearer sense of where the truly intractable disagreements lie and thus where compromise becomes necessary when the debate reaches the point of policy design.

Here we consider a few debatable assumptions about the proponents of accountability reforms, including standardised assessments. These proponents are not necessarily economically conservative. As has been seen above, reforms can be revolutionary in the sense that they undermine a corrupt state. Undoubtedly, the strong link that at least developing countries see between accountability reforms and the World Bank, in the sense that the World Bank has funded much of the research in this area, can lead to suspicions around underlying ideological agendas. Many of the misgivings can be traced back to the World Bank's structural adjustment policies, beginning in the 1980s, which undeniably had harmful and avoidable effects on certain developing countries and their education systems, though, as Carnoy (1995) points out, the World Bank itself has acknowledged many of the mistakes. Clearly countries must engage critically with international organisations such as the World Bank. This is arguably what Brazil has done, which has selectively used World Bank assistance in developing its standardised assessment systems.

The proponents of accountability reforms are not necessarily hostile to teachers, or teacher unions, as will be seen below. Attempts to understand teacher preferences and design appropriate monetary and non-monetary incentives in teaching are an admission of the centrality of teachers in educational reform. Neither do the proponents deny the importance of adequate budgets and resources in schools. Yet in this regard it must be acknowledged that there is an important flashpoint. Economists have argued that at least in developing countries teachers tend to capture a greater share of the education budget than they should, leaving insufficient spending on the non-personnel tools that teachers need to do their job¹⁴. The argument that the accountability reformers intentionally or unintentionally 'dumb down' education by shifting the focus away from less measurable learning towards measurable learning is difficult to defend, at least in developing country context where the point of departure is often extremely poor levels of learning. Improvements in measurable learning in developing countries, where this has occurred, is likely to enhance the ability of pupils to engage with more complex problems. Basic literacy is a prerequisite for any advanced academic pursuits. Economic theory is not compatible with the idea that higher order skills should be suppressed in order to maintain a pliant and exploitable workforce. On the contrary, economic theory tends to value critical thinking and problem solving in the labour force as key factors behind faster economic growth.

The matter of teacher autonomy raises interesting differences. The economic literature promotes greater autonomy for schools and school districts, through decentralisation. This is part of an efficiency argument that says that central authorities are too detached from schools to take decisions relating to, for instance, teacher in-service training. Instead, the funds for the activity should be transferred to a local level. At the same time, the public management side supports centrally designed monitoring systems. Schools should all be subject to the same

¹³ Ibid. pp. 57, 59.

¹⁴ See Pritchett and Filmer (1997).

quality standards. The proponents therefore support a certain mix of local autonomy and central monitoring. What teacher unions are likely to find threatening in this is the decentralisation of decisions around teacher salaries, a shift which is likely to diminish the ability of unions to bargain. Thus whilst the teacher side would favour teacher autonomy, this is unlikely to extend to local autonomy with respect to salary decisions.

If there are inescapable points of conflict in the education policy polemics described in this section, these relate to the determination of teacher pay and the power of teacher unions. Moreover, respecting teacher autonomy where this is valued for professional reasons is a further possible area of contestation. Here, however, it is noteworthy that violations of teacher autonomy are understood largely in terms of an organisational culture that under-values the role of the teacher, as opposed to specific interventions such as standardised assessments. Assessments that respect both external standards and teacher autonomy seem a possibility deserving careful attention.

4 Circumstance, ideology and policy choices

Ideological or quasi-ideological disagreements with respect to standardised assessments appear to emerge with respect to certain combinations of policy issues and national circumstances, which are listed below. The discussion in this section proceeds through an imaginary matrix framed by these two dimensions. The intention is to expand on the discussion in the previous section and relate it to some policy specificities.

| Policy issues (first dimension) | Circumstances (second dimension) |
|---|--|
| Use of sample | Layers of governance |
| Multi-stakeholder planning | Strength of teacher unions |
| Universal assessments | Professionalism amongst teachers |
| ...with publicly available school results | Overall education reform package |
| ...with rewards and sanctions | Community-school accountability structures |
| Explicit link to support | School control over resources |
| | Technical capacity of the authorities |

Assessment programmes such as the international PISA are strictly sample-based and can therefore not be linked to rewards or sanctions in any way, and are unlikely disturb a teacher's work because the probability of one's school being sampled is rather small. Sample-based standardised assessments should therefore be uncontroversial. Yet they can be of great value in initiating change. In South Africa, the policy shift towards improving the quality of education for historically disadvantaged pupils was to a large degree a result of findings from SACMEQ and the country's sample-based national assessment programme. Veloso says the following about Brazil (here all national assessments were sample-based until 2005):

It should be kept in mind ... that the very fact that we are now able to measure the quality of education in Brazil represents a major step forward. Before the mid-1990s, there was no assessment system in the basic education sector capable of monitoring trends in the quality of schooling. The establishment and ongoing improvement of such a system during the last 15 years constitutes a critical education policy innovation that has served as a basis for the introduction of a range of successful interventions. (Veloso, 2010: 14)

It is not clear what kind of misuse of PISA data Education International refers to in its policy paper (see earlier discussion). The only conceivable way in which sample-based systems can be linked to rewards is if increases in national salary scales for teachers are in some way pegged to improvements in learning. Such an approach, though completely in keeping with public management thinking, appears not to have been seriously considered anywhere. It is likely that misgivings about PISA relate not to what PISA itself will be used for, but what

further innovations programmes such as PISA may herald. This suggests that policymakers need to be sensitive about the ‘baggage’ that is perceived to come even with sample-based assessments and should explain to, for instance, teacher unions what the intended larger trajectory for improving education is¹⁵.

Benveniste (2002: 96) explains that poor results from a standardised assessment are more likely to de-legitimise a national government of a more centralist country than a more federalist country. In de-centralised countries it is easier for a national government to deflect criticism towards sub-national governments. This suggests that in centralised countries it is especially important to introduce mechanisms that protect a national assessment system from political interference such as the manipulation or concealment of results.

The World Bank’s ‘textbook’ (Greaney and Kellaghan, 2008) on national assessments emphasises the importance of having a multi-stakeholder national steering committee that provides strategic direction to the national assessment programme to ensure that it serves its intended purpose, which ultimately is to improve learning outcomes. In other words, governments should not proceed entirely on their own. In reality, some governments break this rule yet appear to produce a workable system. Arguably this is true of Brazil, where the federal ministry’s semi-autonomous assessment unit has pursued a relatively centralised approach. The case of Uruguay has been used to illustrate that it is important and possible to include teacher unions in the steering of the national assessment in a context of politically powerful unions. Pedro Ravela (2005, 2006), head of the Uruguayan government’s assessment unit from 1995 to 2001, attributes much of the success in implementing the country’s national assessment and getting it to influence learning positively to the way in which unions were involved. To begin with, unions were suspicious about standardised assessments, for reasons discussed in the previous section. However, close collaboration between the government and unions in the planning of the system resulted, firstly, in familiarisation with the approach on the part of unions, which allayed initial fears and, secondly, the realisation on the part of government that certain compromises would have to be made in order to retain teacher support for the programme. The key compromise was certain controls over access to information would be in place (we discuss this below).

The literature suggests that some careful weighing up needs to occur with respect to the governance of the assessment programme. Wide participation brings certain benefits, yet the technical challenges relating to, in particular, test and sample design, should not be underestimated. If expert participation is not strong enough, comparability across time and space is compromised. If results are not credible, then no amount of multi-stakeholder buy-in will save the programme.

Inevitably, the stakes and political tensions rise when a country moves from a sample-based national assessment to a universal one that covers all pupils (perhaps in a certain grade). This is a step that Brazil took in 2005 and South Africa in 2011. Clearly it is at this point that multi-stakeholder buy-in becomes more critical. Universal testing applied by multiple layers of government in an uncoordinated way can become particularly problematic. In South Africa, a few provincial systems of universal assessment existed before the full implementation of the national system in 2011 and the latter has stimulated further expansion in provincially driven testing. The risks entailed by simultaneous programmes driven by different levels of government are evident in the United States, where standards set by the state and the federal levels clash in ways that teachers find infuriating (Delandshere and Petrosky, 2004: 3). Moreover, assessments driven by multiple levels of government increase the overall volume of assessing. In the United States, complaints from teachers often seem to relate more to the volume of testing than standardised testing per se.

¹⁵ This is the argument made by Benveniste (2002: 108).

Much of the documented opposition to standardised testing is from developed countries such as the United States and the United Kingdom. One explanation for this in the case of the United States seems to be the sheer volume of standardised testing occurring in this country. Another reason would be that teachers in developed countries tend to be better trained and in a better position to critique the pedagogy and consistency of the testing systems. Yet even in developed countries with poorer levels of teacher training, technical weaknesses in the testing system are likely to fuel opposition to standardised assessments. Bruns *et al* (2011: 65) point out that acceptance of assessment systems is sensitive to how feedback is given to schools after testing. No feedback or poorly formatted feedback lessens the credibility of the programme. Ravela (2006: 11) explains that presenting results to schools according to assessment categories that teachers were already very accustomed to, and ensuring that feedback reached schools forty days after testing had occurred, were factors that upheld the status of the assessment system. To a large degree Bruns (2010) attributes the widespread acceptance of Brazil's universal system to the technical competence of the federal government in developing, over a relatively short space of time, procedures for external marking of test scripts from around 175,000 schools and feedback in the form of a school scorecard covering not just test results, but also enrolment patterns. In fact, a key feature of the Brazilian system is the combination of test results and pupil grade survival within a composite index. This design was aimed at ensuring that schools would not 'game' the system by encouraging worse performing pupils to drop out of school (a real risk in the case of Brazil due to problems of poverty and child labour).

Universal testing brings with it important decisions around how to disseminate test results. Here the economists tend to consider maximum dissemination as optimal. More accurate information increases accountability which improves incentives and decision-making. This has indeed been the thrust in many countries. In Brazil and Chile, individual school results have been made downloadable from the websites of the national ministries. However, the solution is not always as straightforward as this. In the case of Uruguay, it has been argued that it was necessary to confine access to school-level results to the school principal, who would then be free to use the information in whatever way was deemed suitable by the school. This was necessary to secure needed teacher union buy-in. Publication of school-level information was considered to be insensitive and unconstructive. Instead, aggregated statistics above the level of the school would be made available publicly and the education administration would use school-level results to target support. Many transparency proponents would consider this a cowardly concession to inefficiency and secrecy. Yet it is noteworthy that the influential 2010 McKinsey report on education observed that Asian countries, whose schools generally perform well, are also reluctant to disseminate school results too freely, largely due to cultural reasons. An education administrator is quoted as saying: 'No good for our students could ever come from making school data public and embarrassing our educators'¹⁶. What this suggests is that publication of school results should occur after careful consideration of political and cultural factors.

5 A look at one country: South Africa

6 Conclusion

¹⁶ Mourshed, Chijioke and Barber (2010: 70).

References

- Benveniste, L. (2002). The political structuration of assessment: Negotiating state power and legitimacy. *Comparative Education Review*, 46(1): 89-118.
- Bishop, J.H. (1997). The effect of national standards and curriculum-based exams on achievement. *The American Economic Review*, 87(2): 260-264.
- Bruns, B. (2010). *Achieving world class education in Brazil: The next agenda*. Washington: World Bank. Available from: <http://www.iepecdg.com.br/uploads/livros/1012achieving_world.pdf> [Accessed October 2011].
- Bruns, B., Filmer, D. & Patrinos, H.A. (2011). *Making schools work: New evidence on accountability reforms*. Washington: World Bank. Available from: <<http://siteresources.worldbank.org/EDUCATION/Resources/278200-1298568319076/makingschoolswork.pdf>> [Accessed October 2011].
- Carnoy, M. (1995). Structural adjustment and the changing face of education. *International Labour Review*, 134(6).
- Carnoy, M. & Rhoten, D. (2002). What does globalization mean for educational change? A comparative approach. *Comparative Education Review*, 46(1): 1-9.
- Castro, A.M. (2008). *Accountability: Uma nova estratégia de controle da gestão escolar*. Available from: <http://www.jurandirsantos.com.br/outros_artigos/ea_accountability_uma_nova_estrategia_de_controle_da_gestao_escolar.pdf> [Accessed October 2011].
- Contreras, D., Flores, L. & Lobato, F. (2003). *Monetary incentives for teachers and school performance: The evidence for Chile*. Santiago: University of Chile. Available from: <http://www.simce.cl/fileadmin/publicaciones-BD-simce/Contreras_paper.pdf> [Accessed October 2011].
- Delandshere, G. & Petrosky, A. (2004). Political rationales and ideological stances of the standards-based reform of teacher education in the US. *Teaching and Teacher Education*, 20: 1-15.
- Education International (2011). *Policy paper on education: Building the future through quality education*. Brussels. Available from: <<http://www.ei-ie.org>> [Accessed October 2011].
- Greaney, V. & Kellaghan, T. (2008). *Assessing national achievement levels in education*. Washington: World Bank. Available from: <http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079993288/assessing_national_achievement_level_Edu.pdf> [Accessed March 2010].
- Gustafsson, M. & Mabogoane, T. (2010). *South Africa's economics of education: A stocktaking and an agenda for the way forward*. Stellenbosch: Stellenbosch University. Available from: <<http://ideas.repec.org/p/sza/wpaper/wpapers105.html>> [Accessed March 2010].
- Hanushek, E.A. & Raymond, M.E. (2005). Does school accountability lead to improved student performance. *Journal of Policy Analysis and Management*, 24(2): 297-327.
- Inbar, D.E. (1996). *Planning for innovation in education*. Paris: IIEP. Available from: <<http://www.unesco.org>> [Accessed December 2006].
- Jacob, B.A. & Levitt, S.D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, 118(3): 843-877.
- Jennings, J. & Rentner, D.S. (2006). Ten big effects of the No Child Left Behind Act on public schools. *Phi Delta Kappan*, 88(2): 110-113.
- Kellaghan, T. & Greaney, V. (2001a). The globalisation of assessment in the 20th century. *Assessment in Education*, 8(1): 87-102.
- Kellaghan, T. & Greaney, V. (2001b). *Using assessment to improve the quality of education*. Paris: UNESCO: IIEP. Available from: <<http://www.unesco.org/iiep>> [Accessed October 2005].

- Loeb, S. & Strunk, K. (2007). *Accountability and local control: Response to incentives with and without authority over resource generation and allocation*. Available from: <http://www.usc.edu/dept/education/cegov/focus/governance/publications/journals/1acct%20and%20control_loeb%20strunk_ed%20finance%20and%20policy.pdf> [Accessed October 2011].
- Mizala, A. & Romaguera, P. (2005). Teachers' salary structures and incentives in Chile. In Emiliana Vegas (ed.), *Incentives to improve teaching: Lessons from Latin America*. Washington: World Bank: 103-150.
- Moloi, M.Q. (2005). *Mathematics achievement in South Africa: A comparison of the official curriculum with pupil performance in the SACMEQ II Project*. Harare: SACMEQ. Available from: <www.sacmeq.org> [Accessed October 2005].
- Mourshed, M., Chijioko, C. & Barber, M. (2010). *How the world's most improved school systems keep getting better*. New York: McKinsey & Company. Available from: <http://ssomckinsey.darbyfilms.com/reports/schools/How-the-Worlds-Most-Improved-School-Systems-Keep-Getting-Better_Download-version_Final.pdf> [Accessed February 2011].
- Piper, B. & Korda, M. (2010). *EGRA Plus: Liberia*. Research Triangle Park: RTI International. Available from: <http://pdf.usaid.gov/pdf_docs/PDACR618.pdf> [Accessed October 2011].
- Pritchett, L. & Filmer, D. (1997). *What education production functions really show: A positive theory of education expenditures*. Washington: World Bank. Available from: <<http://www.worldbank.org>> [Accessed March 2007].
- Psacharopoulos, G. (1996). Economics of education: A research agenda. *Economics of Education Review*, 15(4): 339-344.
- Ravela, P. (2005). A formative approach to national assessments: The case of Uruguay. *Prospects*, 35(1): 21-43.
- Ravela, P. (2006). *Using national assessments to improve teaching... and learning: The experience of UMRE in Uruguay*. Washington: World Bank. Available from: <http://info.worldbank.org/etools/docs/library/235974/D2_cartagena_2006_PRavela_fv_11sep06.pdf> [Accessed November 2009].
- Resnik, J. (2006). International organisations, the "education-economic growth" black box, and the development of world education culture. *Comparative Education Review*, 50(2): 173-195.
- Sen, A. (1999). *Development as freedom*. New York: Alfred A. Knopf.
- Shalem, Y. (2003). Do we have a theory of change? Calling change models to account. *Perspectives in Education*, 21(1): 29-49.
- South Africa: Department of Basic Education (2011). *Report on the Annual National Assessments of 2011*. Pretoria. Available from: <<http://www.education.gov.za>> [Accessed October 2011].
- Spaull, N. (2011). *Primary school performance in Botswana, Mozambique, Namibia and South Africa*. Stellenbosch: University of Stellenbosch. [Forthcoming]
- Stiglitz, J.E., Sen, A. & Fitoussi, J.-P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris Available from: <http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf> [Accessed September 2009].
- Sunderman, G.L., Tracey, C.A., Kim, J. & Orfield, G. (2004). *Listening to teachers: Classroom realities and No Child Left Behind*. Cambridge: Harvard University. Available from: <<http://www.eric.ed.gov/PDFS/ED489176.pdf>> [Accessed October 2011].
- The Economist* (2010). Education in Brazil: No longer bottom of the class. London. December 11 to 17.
- UNESCO (1990). *World Declaration on Education for All*. Jomtien Available from: <http://www.unesco.org/education/efa/ed_for_all/background/jomtien_declaration.shtml> [Accessed March 2009].
- UNESCO (2006). *Education for All global monitoring report 2007: Strong foundations*. Paris Available from: <<http://www.unesco.org>> [Accessed November 2006].

- Veloso, F., Pessoa, S., Henriques, R. & Giambiagi, F., eds. (2009). *Educação básica no Brasil*. Rio de Janeiro: Elsevier.
- Wexler, P. (1982). Ideology and education: From critique to class action. *Interchange*, 13(3).
- World Bank (2010). *Achieving world class education in Brazil: The next agenda*. Washington. Available from:
<http://www.iepecdg.com.br/uploads/livros/1012achieving_world.pdf> [Accessed October 2011].
- Ziliak, S.T. & McCloskey, D.N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice and lives*. Ann Arbor: University of Michigan Press.