ED/GEMR/MRT/2017/P1/13

United Nations
Educational, Scientific and
Cultural Organization

Global
Education
Monitoring
Report

Background paper prepared for the 2017/8 Global Education Monitoring Report

Accountability in education: Meeting our commitments

Towards sustainable, collective and participatory accountability in Education 2030

This paper was commissioned by the Global Education Monitoring Report as background information to assist in drafting the 2017/8 GEM Report, Accountability in education: Meeting our commitments. It has not been edited by the team. The views and opinions expressed in this paper are those of the author(s) and should not be attributed to the Global Education Monitoring Report or to UNESCO. The papers can be cited with the following reference: "Paper commissioned for the 2017/8 Global Education Monitoring Report, Accountability in education: Meeting our commitments". For further information, please contact gemreport@unesco.org.

Dr Radhika Gorur

2017

Abstract

UNESCO has made a commitment to 'inclusive and equitable quality education' and promoting 'lifelong learning opportunities for all'. To ensure the successful attainment of this goal by 2030, it seeks to develop measures for monitoring progress and to regulate the actions of various actors. This paper analyses the assumptions that underpin the strategies outlined in *Education 2030 Framework for Action* (UNESCO, 2015a) and discusses the possible ramifications of the outlined approaches. It cautions against conflating the regulatory and informative aspects of accountability, and encourages a 'formative' rather than 'summative' form of accountability. In keeping with Education 2030's vision of collective, participatory and sustainable education reform, it advocates collective, participatory and sustainable accountability practices that are tolerant of uncertainty and focused on learning and improvement rather than accountability for compliance. Some possibilities for the development of such approaches are identified.

Outline

1. Accountability in UNESCO's Education 2030 agenda

UNESCO's Education 2030 agenda is bold and ambitious. It has made an urgent and serious commitment to '[e]nsure inclusive and equitable quality education and promote lifelong learning opportunities for all'. To achieve this goal, one of the key strategies stressed in *Education 2030 Incheon Declaration and Framework for Action* (UNESCO, 2015a, p. 32) is the need for 'good governance and citizen-led accountability in education'. The words 'accountability', 'accountable' and 'account' (as in holding various actors to account) appear no less than 24 times in the 32 pages of the document.

Education 2030 is informed by the experience and the reviews of the *Millennium Development Goals* (MDGs) and in particular *Education for All* (EFA), and specifically alludes to the lessons learned from EFA. UNESCO has noted with some disappointment that while considerable progress was made globally, not a single country had met all the goals agreed to in EFA. Moreover, even where impressive overall progress was achieved, inequities had not been adequately addressed. Why was this the case? Were the goals appropriate? What new goals are relevant now? What strategies are most likely to work? And how can success be ensured this time around – success for *all*? These questions have been deliberated in a series of meetings, which have concluded that countries had interpreted and prioritised the six EFA goals variously, and were confounded by competing demands, lack of funds and other challenges. Moreover, the six goals in EFA did not interest all nations equally. Some of the EFA goals appeared more compelling or had gained more attention – especially universal primary education – and this had deflected attention from other goals. In other words, the EFA goals themselves competed against each other, rather than supporting each other as might have been visualised (UNESCO, 2013). In response, the Incheon Declaration committed to a 'single, renewed education agenda that is holistic, ambitious and aspirational, leaving no one behind' (p. iii). Over the next 15 years, UNESCO must successfully impose this global agenda on a large and diverse set of governments with complex, idiosyncratic and very particular problems; monitor and regulate these governments and ensure they stay true to their promises; determine whether progress is being made and measure it; and find mechanisms to hold all the parties involved to account.

This is a regulatory project of heroic proportions. The ordering of social worlds and the regulating of agencies and institutions has always been difficult to achieve at a national or even institutional level; the diversity and number of nations involved in *Education 2030* makes this a particularly challenging task. Much has been written about the challenges faced by various similarly ambitious regulatory projects, such as creating national measures for weights (Scott, 1998), or developing international indicators in education (Bottani, 1996; Gorur, 2014a; Sauvageot, 2008). These large-scale efforts to standardise, regulate and render accountable (for example, Gorur, 2015a; Scott, 1998) offer lessons for the post-2015 agenda.

This 'think piece' examines how is accountability being understood and how its operationalization is visualised in *Education 2030.* What assumptions about the way accountability works in various aspects of education – from the school or community level to the level of international governance and collaboration – are evident in these plans? What understandings of education systems and of community participation are inscribed into the Framework? What are some cautions that might be appropriate in light of these plans? Drawing on both theoretical and empirical material, the paper critically examines these questions. It argues that while the Incheon Declaration and the *Framework* emphasise the involvement of the civil society; a participatory approach that involves multiple stakeholders; a contextualised interpretation of the goals and development of strategies; empowerment and capacity building at the grassroots; and other inclusive, holistic and collective approaches aimed at on going learning and improvement, 'accountability' as a concept and practice appear to be visualised in a mostly conservative, top-down and bureaucratic manner. There is much focus on 'evidence-based' measures, where the evidence proposed is not much different to what is already being used in several parts of the world, and which has proven to be ineffective. This paper urges UNESCO to reconsider the purposes of

measurement, and urges a shift in focus to on-going evaluation for improvement rather than measurement for accountability. Indeed, it suggests a re-visioning of accountability altogether: to complement UNESCO's vision of a participatory, collective, sustainable approach to education reform, this paper recommends an accountability regime that is collective, participatory and sustainable.

2. Accountability: a complex concept

Rarely used until a few decades ago, the term 'accountability' has become ubiquitous globally in public administration (R. Mulgan, 2000). It is generally associated with transparency, governance and new public management practices (NPM) which characterise evidence-based policymaking (EBP) (Gorur, 2011c). EBP represents a shift away from the traditional phenomena associated with policy – politics, people and power (Parsons, 2002) – towards a kind of neutral and technocratic process based on evidence, objectivity and rationality (Gorur, 2011c; Gorur & Koyama, 2013). Accountability, with its emphasis on numeric instruments and mechanisms, is key to this rationality.

There is general agreement that 'accountability' is about 'the obligation to give an account of one's actions to someone else, often balanced by a responsibility of that other to seek an account' (Normanton, 1966). Because it involves someone who gives an account to someone who seeks it, it involves both responsibilities and obligations. Bovens (2006, p. 7) defines accountability as 'a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor may face consequences'. Mulgan (2000, p. 555) elaborates the features of this kind of accountability as being: 'external' (an account is given to someone else); involving 'social interaction and exchange' (one party seeks answers, clarification and possibly rectification, while the other responds to these demands); and implying 'rights of authority' of those who demand accountability. Whether the imposition of sanctions is part of accountability depends on whether it is understood as 'giving an account' or 'calling to account'. This distinction can be crucial to the extent to which those giving an account see accountability as having value for their own practice, or merely as an unavoidable obligation that may have no real benefit to them.

2.1 Typologies of accountability

Given its ubiquity, accountability has been much analysed, and a range of typologies have sprung up. Romzek and Dubnick (1987) elaborate four distinct types of accountability within government organsiations: *political, legal, bureaucratic*, and *professional* accountability. A slightly different set is elaborated by Erkkiä (2007), who identifies political, bureaucratic, professional and personal accountability. In addition, he introduces two new types – *performance* (output or client-oriented accountability operationalised through competition and self-regulation with the context of marketised settings, and *deliberation*, characterised by being interactive, deliberative, open and public, and operationalised within the public sphere through debate, deliberation and transparency. Particularly with regard to accountability measures connected with 'teacher quality', this distinction between 'performance' oriented and 'deliberation' oriented accountability is useful. One could imagine how a performance approach to accountability might interfere with a deliberative approach, to the detriment, I would argue, of improvement in the quality of teaching.

Woolgar and Neyland (2013, pp. 31-33) distinguish two types of accountability: *mutual accountability*, 'a two-way process whereby the accountability of government officials to the citizenry makes possible citizens' responsibility for the acts of government'; and *organizational accountability*, 'the processes whereby organizations, for example, are required to produce information on which they are held to account'. Organisational accountability points to the instruments and processes involved in performances of

accountability, such as the adoption of explicit standards that are often "codified in law, administrative regulations, bureaucratic checks and balances, or contractual obligations to other organizations" (Kearns, 1996, p. 66). Standards, benchmarks, contracts and performance measures are thus artefacts in which accountability obligations are inscribed, and which enable the enforcement of these obligations (Koliba, Mills, & Zia, 2011). Examining the *Framework,* Education 2030 appears to seek both types of accountability – that which involves citizens and a range of stakeholders as well as that which is based on benchmarks, indicators and measurement. Norris (2014) distinguishes between *principal-agent accountability* – where someone actively holds another to account, and *reflexive accountability*, in which actors change their own behaviour in light of information of the kind contained in 'monitoring information, peer review, policy evaluations, and policy data from other localities. In education contexts post-bureaucratic practices of 'intimate accounting' (cf Asdal, 2011; Gorur, 2015b) has produced a hybrid that involves both these types: principal-agent accountability involving the governments, for example, holding schools to account, and at the same time, high-stakes accountability that is pushing teachers and schools to engage in versions of reflexive accountability (even if 'information' generated by high-stakes assessments are not always seen by teachers as reliable). Foucauldian scholars would see this as the panopticon turned inwards. 'Intimate accounting' processes enable the development of personalised targets, self-monitoring and regulation, where, rather than governing from a distance, regulatory mechanisms allow regulators to penetrate individual schools and enrol them into monitoring themselves (Gorur, 2015b).

2.2 Accountability as culture

In governance and administration, accountability – and indeed policy making and governance as a whole – are seen as rational, neutral and generally straightforward and objective processes. This conceptualisation is challenged by a range of scholars. Ranson (2003, p. 462) sees accountability as 'a *social practice* pursuing particular purposes, defined by distinctive relationships and evaluative procedures'. Power (1996, 1997) views accountability and the audit processes integral to it as 'rituals of verification' within contemporary 'audit societies' which, he says, are enamoured with accountability mechanisms. Here the rationality of systems of accountability is not seen as self-evident. For Strathern (2003), accountability practices are part of contemporary 'cultures in the making'. This view also disputes rationality and objectivity as something outside of cultural practices and processes – i.e., 'objectivity' and 'rationality' are themselves seen as culturally produced).

2.3 The 'accountability deficit' in education

Most countries today have a system of public education, and education is seen as a public good. With taxpayers' money spent in public education, it is thought that both governments, which are ultimately responsible for education, and schools and universities and other institutions where teaching occurs, should be responsible to the public. An OECD study by Hooge, Burns and Wilkoszewski (2012) suggests that the increased autonomy allocated to schools in recent decades has created a problem for states, which have to find ways to hold these autonomous schools accountable. Their analysis suggests that comparative international assessments such as Trends in Mathematics and Science Studies (TIMSS), Progress in International Reading Literacy Survey (PIRLS) and Programme for International Student Assessment (PISA) have focused attention on national or system-level performance, whereas school-based accountability has promoted decision-making at a local or regional level, and over the past decade or so, this has made accountability a critical issue. They identify two 'accountability deficits': a. there is no mechanism by which schools and school boards can be held responsible to the public for their decisions and performance; b. in its anxiety to hold schools accountable, there is heavy reliance on standardized assessments, which are limited in the range of outcomes they can survey. The limitations of mechanisms and instruments for measuring school performance is the second deficit highlighted by Hodge et al. It is evident from the above that accountability is a complex concept, and when various actors allude to accountability, they may be referring to quite different concepts and practices. This could lead to a lack of clarity on what is desired, and when diverse actors are working collaboratively, this confusion could result in ineffective and wasteful practices. The next section examines how accountability is conceptualised in Education 2030's *Framework*.

3. Conceptualisations of accountability in Education 2030

Accountability is visualised as central to the success of Education 2030, and the *Framework* makes many references to the purpose, concept and practice of accountability and how it will feature in achieving the goals of the program. This section briefly presents some understandings of 'accountability' implicit and embedded in the *Framework*.

3.1 A principle

Accountability is seen as a self-evidently good, basic principle and an essential ingredient of responsible governance within a humanistic vision that honours citizens' rights. For example:

> *[Education 2030] is rights-based and inspired by a humanistic vision of education and development, based on the principles of human rights and dignity, social justice, peace, inclusion and protection, as well as cultural, linguistic and ethnic diversity and shared responsibility and accountability. (UNESCO 2015a, p. 4)*

Here accountability is visualised as an essential ingredient of a just world. Neither imposed nor demanded, it is seen as a feature of shared responsibility and good governance in an ideal society. This articulation inscribes particular understandings of good governance, progress, dignity and social justice that are taken to be universally valued. It evokes order, stability, responsibility and fairness.

3.2 A key to good governance

Education 2030 frames accountability as a key to policy planning, decision-making and rational and systematic governance.

Accountability as data

In the era of 'evidence based policy', good governance requires clear, reliable knowledge. Accountability mechanisms are seen as providing governments with unambiguous, unbiased and clear data that would enable correct decision-making and tidy organization of administrative efforts into well-defined and specific targets, 'goalposts' and benchmarks:

> *Governments are expected to translate global targets into achievable national targets based on their education priorities, national development strategies and plans, the ways their education systems are organized, their institutional capacity and the availability of resources. This requires establishing appropriate intermediate benchmarks (e.g. for 2020 and 2025) through an inclusive process, with full transparency and accountability, engaging all partners so there is country ownership and common understanding. Intermediate benchmarks can be set for each target to serve as quantitative goalposts for review of global progress vis à vis the longer-term goals. (UNESCO, 2015a, p.10)*

Accountability as a method of increasing (school) responsibility

The involvement of the media and the publication of data such that it can be widely accessed on public websites, in easily understood comparisons, rankings, tables and graphs, is a way to create 'informed publics' who can then participate by exerting market pressures on education institutions (Gorur and Koyama, 2013; Gorur, 2014). The *Framework* also visualises informed publics holding the government to account:

> Partnerships at all levels must be guided by the principles of open, inclusive and participatory policy dialogue, along with mutual accountability, transparency and synergy. Participation must begin with the involvement of families and communities to boost transparency and to guarantee good governance in the education administration. *Increased responsibility at the school level could strengthen efficiency in the delivery of services. (UNESCO, 2015a, p. 24)*

Accountability as a coordinating mechanism and a regulator of behaviour

In a recent paper (Gorur, 2015b), I have theorized how such accountability mechanisms as Australia's My School website which publishes comparative performance data of every school in the country, as an 'interesting object' – a policy object which gathers a variety of actors together, bringing some actors unexpectedly into new relations with each other. In this way, it becomes a coordinating mechanism with articulates new relations – for example, parents may become allies of the government in keeping watch over schools and teachers. In the *Framework*, accountability is visualized as the thread that connects various levels of actors in a cascading, hierarchical pattern:

> *It is imperative for all partners to embrace the common vision of Education 2030 outlined in this Framework for Action and to be held accountable: multilateral organizations should be accountable to their Member States, education ministries and other related ministries to citizens, donors to national governments and citizens, and schools and teachers to the education community and, more broadly, to citizens. (UNESCO, 2015a, p. 25)*

Since a major issue identified with regard to EFA was that at various levels, actors did not do what they had promised to do or were required to do, the idea of accountability as a regulator of behaviour is implicit in Education 2030. It is seen as means of ensuring that actors at various levels comply.

3.3 A regulator of governments

An essential value embedded in Education 2030 is that of participatory democracy, and the protection of human rights through mechanisms by which citizens can actively become involved in the common goal of ensuring equitable and quality education for all. A key aspect to this is the ability of citizens, NGOs and other organisations to have clear information about government decisions and to be able to hold governments to account. Here a 'bottom-up' as well as lateral accountability is visualised.

Accountability as a facilitator of participatory governance

Informed publics can not only hold education institutions to account, but also exert pressure on governments to achieve the targets they've published and fulfil the promises they've made:

> *[D]ocument and share evidence from practice, from citizens' assessments and from research to inform structured policy dialogue, holding governments accountable for delivery, tracking progress, undertaking evidence-based advocacy, scrutinizing spending and ensuring transparency in education governance and budgeting. (UNESCO, 2015a, p. 24)*

Accountability as a scaled regulatory mechanism

Accountability is visualised at various levels, from local, through national, regional and global. Accountability at a global level, to be instituted by GEMR, will be achieved in part through the establishment of global education indicators.

> *The GEM Report will be the mechanism for monitoring and reporting on SDG 4 and on education in the other SDGs, with due regard to the global mechanism to be established to monitor and review the implementation of the 2030 Agenda for Sustainable Development. It will also report on the implementation of national and international strategies to help hold all relevant partners to account for their commitments as part of the overall SDG follow-up and review. (UNESCO, 2015a, pp. 28-29).*

3.4 A strategy for improvement

Ultimately, accountability is not seen as an end in itself – it is seen as key to improving education outcomes and achieving the Education 2030 goal of equitable, quality education for all. Implicit here is the idea that clear data are important for governments to take proper, informed decisions that will regulate education and lead to improved effectiveness in the deployment of funds, in the regulation of education institutions, in the learning

that occurs, in the quality of teaching and in the alignment of learning with desired national education goals. Indeed, improvement in education is tied to improved equity, particularly with regard to gender, a reduction in poverty, improved health and nutrition, and even a reduction in conflict and increased security across a range of fields. More specifically, accountability is seen as improving the outcomes of education by strengthening the system and improving its efficiency, ultimately resulting in improved outcomes for citizens and nations.

Strengthening the system

Accountability is seen as strategy for improving education and ensuring quality, and a way of strengthening education systems.

> *[T]o ensure quality education and conditions for effective education outcomes, governments should strengthen education systems by instituting and improving appropriate, effective and inclusive governance and accountability mechanisms... (UNESCO, 2015a, p.8)*

Accountability as an efficiency mechanism

Accountability is also seen as an essential ingredient in good management and governance:

> *Improving governance and accountability can increase efficiency and effective use of existing resources and ensure that financing reaches the classroom. (UNESCO 2015a, p.30)*

From the above, it is clear that accountability is being conceptualised in a variety of ways in the *Framework*. It is seen as a principle, a strategy, a solution, and a coordinating mechanism. It is seen as operating at global, national and quite local scales. It is a mechanism that enables administration in a rational, ordered way, with decisions based on objective information. Both top-down or organisational accountability, as well as mutual accountability are visualised. These understandings are implicit rather than explicit in the *Framework* – and this may mean that they are prone to even more interpretations as various nations, and various stakeholders within each nation, operationalize the *Framework*.

Interestingly, few, if any, references are made in the *Framework* to the very well documented critical literature in education and policy studies on particular forms of accountability, including the unintended negative consequences of accountability that are occurring in education in contemporary times (see, for example, Lingard, Martino, Rezai-Rashti, & Sellar, 2016; Meyer & Benavot, 2013). The next section looks at some complications and issues with accountability that are important to consider and to anticipate when promoting accountability in Education 2030.

4. The challenges of accountability for Education 2030

Given the extremely challenging and ambitious task ahead for Education 2030, it is important to anticipate some issues it may confront with regard to accountability. While the scale and scope of accountability conceptualised in Education 2030 are unprecedented, there are examples of strong accountability efforts at the national and transnational levels that could serve as cautionary tales and empirical experiments from which Education 2030 could well draw useful lessons. In this section, these examples are drawn upon to anticipate the challenges ahead for Education 2030.

4.1 Policy making is complex and chaotic

Straightforward accounts of accountability as part of the rational, transparent and apolitical approach to policy deny the complexity, and the multi-scalar, multi-faceted nature of policymaking, particularly in these networked, globalised times (Gorur, 2011c). Rational accounts of policy as proceeding from problem identification to a call for solutions and then implementation and review are just not how policymaking occurs (Gorur, 2011b). In

visualising the nature and role of accountability in Education 2030, it is important to explore how accountability participates in policy assemblages at national and provincial levels.

In a set of interviews I conducted in 2008-09, 18 policymakers at national and international levels (purposively chosen for their senior roles and deep knowledge of the OECD, PISA or experience in the Australian government) provided descriptions of how education policy worked. These lively descriptions presented a picture of dynamic circumstances requiring constant vigilance and response. Far from a stable, regulated world of certainties based on data and projections informing decision-making, my interviewees described policymakers as living in a world of constant uncertainties; of intervening strategically at the right moment; of the influence of particular individuals or ideas; and of the importance of creating a climate for ideas to get accepted as 'fact' or to be absorbed into practice' (Gorur, 2011b). One interviewee explained:

> *You see the theorists draw their little models and so on, [but policymaking] doesn't work like that. It's totally chaotic. And often major reforms are driven by particular ideas that particular people have and particular circumstances. But what does count is the environment ... I think the idea of climate is really important and that's where research plays a role, but it is a very slow, drip-feed type of role. You start changing the ways that people think about the world, and then there'll be some catalyst at some stage and things happen. (Interview Transcript, policy expert, cited in Gorur, 2011b, p. 612)*

A former minister's account further attested to the complexity and unpredictability of policy work:

> *[T]he political activist and the decision maker has to understand how the institutions allocate influence and how the various interests accumulate political resources and how they are going to deploy those in the process and what your role as the authoritative decision maker can be in influencing that outcome ... It all fits, I suppose, in a kind of linear model eventually, although ... very often it is not a rational process which leads to an outcome. That again depends on where you see rationality lie, because there is a political rationality as well. (Interview Transcript, policy expert, cited in Gorur, 2011b, p. 612)*

In my study of PISA and of Australia's 'Education Revolution', I identified such activities as articulating and rearticulating the policy problem, forging alignments, rendering technical (or anti-political), authorising knowledge and reassembling as some of the activities involved in pushing reforms to take root and flourish. These studies also highlighted the role of the mundane, such as routinized activities and the work of various devices, instruments and objects that mediate relations, extend and make durable the reach of humans, translate the interests of actors and render phenomena mobile and transportable (Gorur, 2011a; Woolgar & Neyland, 2013).

For Education 2030, this may mean that beyond each country interpreting the universal goal and committing to specific benchmarks to be achieved, it would be important, at local sites, to explore how the political climate can be influenced to promote the chances of success in meeting the goals. When policymaking itself is understood as a networked, multiscalar activity rather than a tidy process from problem identification to evaluation, the role of accountability in policymaking – and indeed the practices of accountability – need to be conceptualised quite differently.

5. Compliance cultures are not sustainable

Accountability is centrally about trust. Trust is delegated from individuals to calculations. Standards, benchmarks, rankings and other forms of data are seen to arise from the collective integrity and expertise of anonymous, distant, disinterested and apolitical actors (Gorur, 2011a, 2013; Gorur & Koyama, 2013; Porter, 1995), rather than proximate, passionate and invested actors such as teachers. Rather than pin faith on the integrity or wisdom of individuals, trust is delegated to these instruments that may be seen as more dependable, unbiased and

honest. Calculations act as 'anti-political' instruments which 'limit the possibility of disagreement' (Barry, 2002, p. 272).

Analysing the relationship between 'accountability' and 'responsibility', Mulgan (2000, p. 558) concludes that 'the expansion of 'accountability' has … been accompanied by a corresponding contraction of 'responsibility', with the latter being increasingly confined to issues of 'personal culpability, morality and professional ethics'. Accountability could also be seen as displacing responsibility – for example, from those responsible for curriculum, funding, the organisation and administration of schools and so on, to teachers and schools.

Romzek and Dubnick (1987, p. 228) argue that the conceptualisation of accountability as 'answerability', evokes a sense of 'limited, direct, and mostly formalistic responses to demands generated by specific institutions or groups in the public agency's task environment' and does not adequately describe how accountability mechanisms are encountered and negotiated on a day-to-day basis in institutions. They advocate a broader conception: 'public administration accountability involves the means by which public agencies and their workers manage the diverse expectations generated within and outside the organization'. They contend that government agencies negotiate multiple accountability systems – both internal and external – and this comes at a cost. In their study of NASA's Challenger tragedy, they found that attempts to satisfy political and bureaucratic accountability deflected attention from NASA's core strengths of professional standards and accountability mechanisms, resulting in the disaster. They conclude that increasing certain requirements of accountability can hamstring institutions and create dilemmas for them.

This idea can be related to the distinction made between internal and external accountability made by the British Columbia Teachers' Federation.[i] They suggest that by imposing external measures of accountability, teachers' internal accountability – i.e., their professional code of ethics and their personal ethics – are discounted. While they accept that it should be possible to ask teachers to explain their decisions, the imposition of external accountability shows a lack of trust in teachers and in the profession. Moreover, they argue that external accountability mechanisms are blunt and can only provide distorted views of teaching and learning, as well as leading to well-known ills such as teaching to the test and narrowing of the curriculum.

As with intrinsic and extrinsic motivation, imposing accountability from the outside is likely to produce short-term compliance at best, to the detriment of sustained interest in improvement, self-monitoring and self-regulation. A serious unwelcome side effect with external accountability is the deprofessionalisation and devaluing of teachers. In Australia, for example, it is reported that anywhere from 20 to 45 per cent of trained teachers leave the profession in 2-5 years (Buchanan et al., 2013). Those eligible to take up positions of principals are not taking them up, in part because accountability measures make the task of the principal onerous and stressful.

With the devolution of responsibility to schools or school districts, much accountability in education has focused on central or state governments holding schools accountable for student outcomes using a range of measurement mechanisms and rewards and punishments associated with student performance on state-imposed standardized assessments. The term 'high-stakes' is becoming widely used in this connection. The most common forms of such consequences are in evidence in the US, where, under the *No Child Left Behind* (NCLB) Act and other similar policies, schools that did not meet the required benchmarks faced funding cuts and even closure or amalgamation (Gorur & Koyama, 2013). There are several examples of high stakes accountability mechanisms leading to the narrowing of the curriculum, the exacerbation of inequities, the deprofessionalisation of educators and even to systematic cheating and falsification of test results (Jacob, 2005; Nichols & Berliner, 2007). Thus the very instrument that is designed to keep practitioners and policymakers honest may in fact lead to malpractice; and, more importantly, the bid to improve outcomes for students might serve to work against them.

Some of the consequences attached to poor student outcomes have had particularly perverse effects in education. Student outcomes are dependent on a range of factors, many of which are outside the control of teachers, such as students' home backgrounds, the curriculum adopted and the adequacy, prioritisation and distribution of resources. These factors are usually classified as 'input measures', but they are, in fact, related to expected outcomes from other actors in the 'education collective'. Measurement and accountability should not, therefore, be based only on the 'end outcome' of student performance and equity measures – it should also include such aspects of the environment as may influence the attainment of these goals, such as improved health, the empowerment of women, better employment prospects, reduction in conflict and improved security, improved early pre- and post-natal health and so on. In this regard, UNESCO might be able to work with other UN organisations such as UNDP and UNHCR to set up collectives that would work out which environmental factors required enhancing in particular locations, and then develop comprehensive benchmarks and goals that encompassed the entire collective that is required to work together to improve learning outcomes.

Education 2030 visualises a collaborative, participatory approach among a range of stakeholders toward a common goal: equitable, quality education for all, as interpreted in that country/location. This participatory and collective approach is at odds with relations of accountability of the type described in the *Framework*, and certainly is not in evidence in current cultures of accountability in any social policy sphere. The challenge for Education 2030 would be to figure out how to stimulate and harness not only teachers', but also all involved actors' personal and professional ethics, aligned to the common vision that has been articulated, in such a way that there is minimal need for strong, obtrusive political and bureaucratic accountability. Instead, it would need to get teachers, parents, school administrators, policy makers, academics and education experts as well as professionals in all related social policy domains focused fully on meeting the common goal by identifying needs and working together to address them, without the distraction of being held accountable. That is, the desired behaviour needs to be internalised so that external threats or need for organisational and top-down regulation is minimal. The current model visualised is of a cascade of accountability relations, without distinctions between different types of accountability, and without due attention to how these might interact with each other.

5.1 The leap from 'data' to policy can be treacherous

Accountability refers to the processes by which actors, events and phenomena are made account*able,* i.e., rendered into formats that 'provide an account' (Gorur, 2011c). Thus accountability involves the translation of phenomena into formats that tell a story or paint a picture. In education, numbers such as those generated by the OECD through PISA have been extremely influential internationally (Breakspeare, 2012). National level data have long been generated in a range of countries, and in recent times, these measurements have become much more complex and comprehensive. The accountability measures associated with the US policy No Child Left Behind have come to be widely reported. In Australia, state level data have been replaced by national measures through the National Assessment Program – Literacy and Numeracy (NAPLAN) and the 'like-school' comparisons based on NAPLAN performance. These accounts are expected to produce reliable data that can inform policymaking, and counter any temptation to be ruled by ideology or individual intuition or wisdom. As Mulgan noted, policymaking is increasingly relying on a particular type of data not only for information, but also to reassure the public of the reliability of its decision-making:

> *Governments have become ravenous for information and evidence. A few may still rely on gut instincts, astrological charts or yesterday's focus groups. But most recognise that their success – in the sense of achieving objectives and retaining the confidence of the public – now depends on much more systematic use of knowledge than it did in the past. (G. Mulgan, 2005, p. 215)*

Similarly, Gary Banks noted how evidence can counter unfounded understandings, and that unless evidence was used, the complexities of social phenomena and the uncertainties of reform might go unrecognised, resulting in poor policies:

> *Without evidence, policy-makers must fall back on intuition, ideology, or conventional wisdom—or, at best, theory alone. And many policy decisions have indeed been made in those ways. But the resulting policies can go seriously astray, given the complexities and interdependencies in our society and economy, and the unpredictability of people's reactions to change. (Banks, 2009, p. 110)*

However, many studies have shown that the type of data that are being generated at international and national levels in education today are also problematic. Several issues are highlighted:

Generating internationally comparable indicators is extremely complex

The Technical Advisory Group's (TAG's) proposal, contained in *Thematic Indicators for Monitoring the Education 2030 Agenda* (UNESCO, 2015c), explicitly advocates a shift from measures of enrolment and other indicators of access to measures of performance. Whilst cautioning that not all indicators are best captured at a global level, and that some phenomena are better measured at a regional or national level, TAG suggests that learning outcomes must be measured using global indicators:

> *Indicators for global monitoring must emphasise this renewed focus on outcome measures. The TAG proposes indicators that enable the measurement and comparison of learning outcomes at all levels of education. (UNESCO, 2015c, p. 4)*

In my study of the development of international education indicators in the production of the OECD's *Education at a Glance*, I have traced how a project that started hesitantly, unsure that such comparative indicators could be generated at all, ended up with the routine, annual production of a 550-page document containing a range of indicators with comparative data on dozens of countries. Such indicators were initially expected to be 'used less for narrow, immediate management purposes and more for global analysis of general trends and differences among countries' (Bottani, 1994, p. 335). However, the policy burden that is now placed on such indicators has increased tremendously. Bottani, who headed the OECD's education indicator development program, contends that the 2000s represented a decade of 'obsession with performance indicators and the triumph of comparative psychometry' which brought in several changes to the indicators project of the OECD, including 'improvements in the quality of data, a modification of the theoretical framework underpinning the indicators, and, perhaps the most significant of all, the launch of PISA, reflecting the rising importance of student attainment indicators' (Gorur, 2014a, p. 587). Bottani reports:

> *The most striking result of this phase is the increased importance given to performance indicators. There were 9 of them in the 2001 edition where a distinction was still made between a section dealing with 'performance at individual and social levels, and on the labour market' (5 indicators) and student attainment (4); in the 2002 edition, they increased to 14; 15 in 2003; and 12 in 2004. With the 2002 edition, the section concerning the set of 'performance' indicators which was the last item in the Education at a Glance index for some ten years, was moved up to take position as the first item. (Bottani, 2008, p. 17)*

The point Bottani makes is that the routine use of these data belie the hesitations and the caveats which attended their birth. Now confidently produced, the ability of these indicators to bear the epistemological burden that is placed upon them is no longer questioned. They have become 'matters of fact' (Latour, 2004) and are taken as self-evident and obvious. For Bottani, this represented the triumph of politics over science.

International comparisons can be misleading

Even as statisticians place caveats and provide cautions, international assessments are used to make policy inferences that are often misplaced. A study of PISA data relating to Australia has demonstrated how misleading analysis based on 'average performance scores' (which is the basis for country rankings) can lead to poor policy decisions (Gorur & Wu, 2015). Analysing the data in terms of geographical unit, performance at item level and

student motivation, the study explored how these analyses pointed to very different policy implications when examined in greater detail, than those suggested by the average scores and rankings.

Just as challenging, if not more, is the measurement of equity in an internationally comparable format. There is a very wide variety in the types of inequities faced in different countries and by groups within countries. Any internationally comparative measures would, perforce, ignore issues that are particular to certain groups and countries. In a study of equity measurement in PISA (Gorur, 2014b), I have explored the fragility of the indicators used to effect comparisons of equity, and demonstrated that a small change in the indicators used to measure equity can change the status of a nation from 'low equity' to 'high equity', without any appreciable change in the actual situation itself.

Indeed, what constitutes 'equity' and how it is understood is also important to explore, and is not to be taken as self-evident. Often, within the same country, several understandings of equity are evident, some of them at odds with each other. Equity in terms of equality may be at odds with positive discrimination, which sees equity as a matter of fairness, for example (Gorur, 2014b). Equity is a strong focus in Education 2030, and monitoring progress on equity also features as important in the *Framework*. If equity is to be addressed seriously, very careful, locally nuanced work is required. Extreme care is needed when using international indicators to measure performance or equity. Even in countries such as Australia that have high-quality statisticians who are aware of the limitations of such data, simplistic understandings continue to be used in policy. Explaining the popularity of international rankings in policy, one PISA expert explained:

> *... I think that the complexity of the data, because of all the effort that is put into making sure that things are done properly, makes it quite inaccessible. You can't just grab the data set and chuck it into SPSS and press return and get a result. Rather than draw upon the big resource that is there, they [governments] tend to use the data that is the already digested and interpreted version that the OECD or occasionally some other people put out. There is really a very valuable resource to engage with, if they would take the time. (Interview transcript, PISA expert, cited in Gorur, 2011c)*

In countries where national data and statistical expertise are limited, the chances of relying on international comparisons and arriving at poor policy decisions are very high, and Education 2030 must find ways to contain this.

The influence of PISA on participating systems, and its effects on education policies and student outcomes would provide valuable lessons for Education 2030. Extensive commentary on the issues with drawing policy lessons from such indicators is widely available. International comparisons are inherently reductive. Whilst this is not, in itself, necessarily a problem, the ways in which such comparisons are leading to simplistic policy lessons (for example, the claims made in Barber & Mourshed, 2007), and limiting our policy imaginations (Gorur, 2015a) is well-established. The possible, irreversible damage caused by such 'monoculture' in our education to our social worlds is now beginning to be discussed (Gorur, 2015c). These must serve as cautionary tales for Education 2030.

5.2 Conflating accountability with improvement could backfire

Quite frequently, the same measures that are introduced for accountability are also used to inform efforts to improve education outcomes. In Australia, for example, the National Assessment Program – Literacy and Numeracy (NAPLAN), which is conducted annually on the full cohort of students in Grades 3, 5, 7 and 9, is used as the basis with which to compare 'like schools' – and these comparisons are published on a website accessible to parents and the general public (Gorur, 2015b). The idea is that parents can use these comparisons to exert pressure on schools to perform better. At the same time, teachers and schools are encouraged to use the data generated by the tests to improve their teaching – many training sessions have been run on how to use these

data to diagnose learning and to inform teaching. Thus the accountability and improvement aspects of NAPLAN have got conflated.

This can be – and has been – problematic. In general, if teachers and schools are looking for improvement, it would be in their interest to explore and expose vulnerabilities and areas of deficit so that appropriate efforts are mobilised and funding and support for improvement can be sought. From this perspective, the function of the tests is to reveal areas of improvement – a successful test would be one that yields useful information, particularly about areas of deficit. However, since NAPLAN is also used to judge teachers and schools, there is far more emphasis on getting a *good score* than in gaining *good information*. The pressure of such judgement has resulted in schools and teachers trying to gain better scores by devoting excessive amounts of time in training students for the test and in some cases cheating to gain better scores. This defeats both purposes of the tests.

Parents have also misunderstood the purpose of NAPLAN. Instead of treating these tests as information about their child's learning, many have focused on improving their child's score on the test. Some parents engage NAPLAN tutors or buy NAPLAN practice workbooks for their children. A whole industry supports the raising of student NAPLAN scores. These practices reduce the usefulness of NAPLAN data as *information for improvement*, and incidentally also corrupt them and reduce their use as accountability measures. Other unintended negative consequences have also been reported – including students getting unduly stressed during NAPLAN tests, as teachers emphasise the importance of doing well in the tests. So concerned were various actors about these tests, that a Senate Enquiry was set up to explore the effects of NAPLAN and make recommendations for mitigating against the worst effects.

Sustaining the argument that measurements such as those generated by NAPLAN or the tests that resulted during NCLB are useful in improving outcomes also becomes difficult. NAPLAN results have shown little appreciable improvement in student performance over the seven years since its introduction in 2008 (Smith, 2015). Instead, during this time, Australia's performance on PISA has declined in real as well as relative terms (Thomson, De Bortoli, & Buckley, 2013). In the US, the very costly efforts to raise student performance with stringent, high-stakes testing regime, NCLB, has been tracked from 2003 to 2015. Results show that achievement levels did rise slowly over time. However, the achievement gap between richer and poorer students rose over this period. The achievement gaps for some groups who have been historically disadvantaged increased for some groups and decreased for others.[ii]

If Education 2030 is focused on improving outcomes for students, then relying on measures of performance at national or international levels that are also used for accountability is not adequate – indeed, these measures may backfire. Where evidence has been successfully used to generate improvement, it has often been generated collegially within schools by teachers themselves, often with the help of specialists focused not on pointing fingers at teachers, but on student learning.

A challenge for Education 2030 is how to develop measures that can monitor progress at the system level without alienating schools and teachers, and at the same time, generate and encourage the use of sound evidence that informs teaching. One possibility to consider is to make system-level tests sample-based rather than whole-cohort based, so that there is no temptation to use the large-scale measures to track the progress of individual students, and to then link these numbers to the quality of teachers and schools. A great deal of work is required in developing cultures of collegial approaches to using good assessment data with a view to improving outcomes. Particularly where there is a shortage of well-qualified teachers, this may be quite challenging. But conflating the goals of accountability and improvement may negatively impact the achievement of both goals.

5.3 Isolating factors ignores the inter-relatedness of collectives

TAG signals a shift in focus from 'inputs' to 'outcomes'. This move strongly echoes the story of the OECD's changing focus in the late 1990s, resulting in the development of PISA (Gorur, 2011a), and leading to the current, unwelcome set of measures that almost exclusively focus on 'teacher quality'. It is critical to examine the logic that underpins this shift. A shift from inputs to outcomes signals a much closer and more straightforward relationship between what happens in classrooms and students' performance. Even though the effects of poverty, health, conflict, socio-economic disadvantage, race and ethnicity, cultural disadvantage and a host of other factors on learning and performance are well known, the shift to 'outcomes' ignores these factors. Instead, attention is fixed on 'teacher quality' that is linked in a straightforward way to student performance.

Importantly, it is not easy to distinguish between 'inputs' and 'outcomes' once the roles of various participants is fully recognised. Deploying adequate funding is critical in education. It may be an 'input' as far as the school or student is concerned, but it is also an 'outcome' for the role that is charged with ensuring that adequate funds are ear-marked, disbursed promptly and deployed efficiently. Similarly, a high-quality, relevant curriculum that prepares students adequately not only in literacy and numeracy, but also in citizenship, ethical behaviour and so on can be viewed as an input from the point of view of students and schools, but is an output for the agency charged with producing that curriculum.

The notion of 'teacher quality' is closely linked to that of 'best practices'. The idea of 'best practices' that is focused on classrooms and teachers discounts the complex inter-relationships that affect teaching and learning. Current trends in education measurement, where these variables are isolated and controlled for in calculations, are not very effective in elaborating how and to what extent these factors interact and affect performance. On the contrary, measurement is focused more and more on isolating the effect of schools and teachers, as with Australia's like-school comparisons.

John Hattie[iii] has famously likened the idea of looking for answers to school improvement in structures, class-size, curriculum reform etc. to a drunk searching under the lamppost for a wallet lost in the bushes, because that is where there is more light. The solution, he insists, lies with the teacher:

> *Interventions at the structural, home, policy, or school level is like searching for your wallet which you lost in the bushes, under the lamppost because that is where there is light. The answer lies elsewhere – it lies in the person who gently closes the classroom door and performs the teaching act –the person who puts into place the end effects of so many policies, who interprets these policies, and who is alone with students during their 15,000 hours of schooling. (Hattie, 2003, p. 3)*

The issue with this argument is that when 'the person … gently shuts the classroom door and performs the teaching act', he or she is not 'alone with the students'. The teacher cannot shut out the influence of poverty, hunger, ill-health, conflict etc. and counter the effects of poor policies, an impoverished curriculum – the world is not waiting obligingly outside the door, but is already present in the classroom, influencing student performance. The challenge for Education 2030 would be to first understand how these varied factors interact with each other to influence performance as a collective, and to develop measures of accountability at the level of the collective. Even more importantly, it would need to enable the collective itself to develop measures of internal accountability. Focusing only on 'teacher quality' as a means of improving education outcomes, especially when there is such a shortage of qualified teachers in many countries, will only waste time and misplace energies and resources. The complexity and local specificity in the ways in which various factors affect a particular set of students would require the development of new metrics that are local, in-depth and small-scale, rather than large-scale and standardised across nations or even within a nation.

5.4 Measurement is not just descriptive, but productive

Accountability measures are often critiqued for their inability to represent phenomena accurately or comprehensively. However, reductionism or distortion are inevitable in all forms of research, and arguably this can be factored into training programs in data use to minimise overconfidence in numerical data. Less difficult to guard against is the performativity or ontological politics of numbers. Numbers do no merely represent realities that already exist; they also shape, influence and even produce new realities. This can be elaborated in three ways:

- As soon as something is measured, it provides a vocabulary and a context to speak about the world in a way that was not possible before and this may set in motion actions that change the world. PISA rankings have led to widespread changes in many systems, including changes to school structures, curricula and policies.
- In anticipation of measurement, people may change their behavior. Knowing you will be weighed may cause people to lose weight, knowing that an odometer reader is ahead causes people to reduce speed, wearing a Fitbit makes one walk more. Knowing that their performance will be measured causes students to study for their exams, schools to prepare students for exams etc. Again, this illustrates that measurement actively sets changes into motion.
- For accountability to become feasible at all, a range of actions is required to be undertaken. Indicators must be developed and negotiated, people trained in using them, phenomena manipulated to become 'research sites' or survey data – in short, the character of calculability must be imposed on the phenomenon before it can be measured and rendered 'accountable'.

In focusing on the accuracy and validity of accountability measures, the ontological politics of such interventions is sometimes neglected. I have argued elsewhere (Gorur, 2015c) that the changes that are brought about by accountability measures, which are focused on ordering and regulating societies and are generally intolerant of diversity, can, in the long run, result in irreversible loss of diversity and the establishment of hegemonies. Given that accountability is a productive and world-changing exercise rather than merely a representational one, conversations around accountability need to shift from accuracy, validity, usefulness and adequacy, to issues of responsibility and consequence.

6. Towards sustainable, participatory and collective accountability

Education 2030's vision of purposeful, collective and sustained efforts to effect improvement in education across all sectors in all countries visualises a range of actors working together with a sense of shared responsibility and mutual trust. In its determination and anxiety to ensure that the ambitious goals are met across the board, however, it envisages an accountability regime that is rather narrowly focused, conservative, and intent more on holding people to account rather than effecting improvement. It expects that holding people to account will result in improvement. The *Global Monitoring Report 2015* states:

> *The lessons are clear. New education targets must be specific, relevant and measurable….The future agenda will … need ever-stronger monitoring efforts, including data collection, analysis and dissemination, to hold all stakeholders to account. (UNESCO, 2015b, pp. i-ii)*

Here accountability is visualised as a form of policing and regulation rather than a strategy for on-going, sustained, participatory efforts to improve education. It is seen as a 'summative' activity – where judgement can be passed and blame apportioned, rather than a 'formative' activity that is woven into the reform effort

throughout to get on-going feedback that is fed into the next stage of improvement, in an evolving process. It assumes that various actors know what needs to be done to effect improvement, but they may not do what is required if they are not held to account. This assumption requires examination and may not hold up under scrutiny.

One danger of such assumptions is that accountability becomes an end in itself – a stand-alone activity that is performed by a separate, independent agency. This paper argues for a different understanding of accountability – one where 'accountability' is implicit and integral to improvement, rather than a separate activity. In keeping with Education 2030's vision of involving a range of actors and taking a multi-pronged approach to improving education, it encourages an approach to accountability that is collective, participatory and sustainable, based on the principles detailed below.

6.1 Develop a 'formative' approach to accountability

A shift is required from an external, summative, regulatory approach to accountability, to one that takes a formative approach. The formative approach does not focus on apportioning blame, because accountability is seen as part of the search for on-going improvement. If something does not go according to plan, the response is to try and figure out why and to come up with a different strategy to reach the goal. This way, local good practices, rather than universal best practices, can be discovered and experimented with. A formative and evolving approach would encourage the discovery and use of appropriate practices locally, rather than the identification and imposition of imported and generic 'best practices'. Critics of efforts to identify 'best practices' through comparison argue that:

- What works in one context may not work in the same way in another
- Just because the outcomes of a particular system are good does not mean that the teaching practices are good – we have no way of knowing how much better that system might do with a different set of practices
- The same practices might well be in place in systems that are not doing as well – with less impressive results

A 'summative', 'holding to account' approach assumes that the processes and practices to be adopted for improving education are already known and follow a predictable path, so that benchmarks can be established ahead of time, and as long as teachers and others comply, a predictable improvement and the achievement of predetermined benchmarks is effected. It also assumes that accountability measures would provide the necessary motivation for actors to do what is required, and, as a corollary, that certain required actions are not being performed because of a lack of accountability. This assumption is both unfounded and likely to reduce internal accountability. At best, it is likely to produce a 'tick box' kind of accountability.

Another very important issue with 'summative' forms of accountability is that processes themselves are never questioned, and any unforeseen or unintended consequences of pre-determined processes do not come to be questioned. Where reforms are sponsored by distant others, as will be the case with many countries in Education 2030, external, impartial teams are sent out to assess progress. Such teams seldom have an understanding of the wider ramifications of the reforms, the longer term changes and the ripple effects that may accrue from the changes. They assess progress against a checklist of intended consequences and benchmarks, which may produce a very distorted and impoverished account of the efforts.

A formative approach needs to be developed. This would involve the identification of a dedicated team at each site, invested in the reform effort and armed with a clear agenda, being focused on assessing the situation on an on-going basis and responding to the information obtained as part of the regular practice of improvement. Such

an approach would create a climate of on-going, sustainable accountability that is built into everyday practices, where 'mistakes' or problems or lack of success are not negative outcomes to be punished, but useful information for building better strategies.

Accountability as 'policing' to ensure compliance is not only doomed to fail, succeeding might trigger serious negative consequences in the long term (Scott, 1998). Instead, a focus on less heroic and more local accountability measures focused on more informed and responsible participation on the part of the actors involved is required.

6.2 Focus on collective responsibility rather than individual accountability

Efforts to improve teaching and learning have been a focus of research for decades. Classrooms have been studied by a variety of scholars, including psychologists, sociologists, assessment experts, curriculum specialists, experts in pedagogy and even architects and economists. Researchers have studied classrooms using a raft of methodologies, theoretical frameworks and disciplinary lenses. Classrooms have been subjected to qualitative and quantitative analyses, longitudinal studies, ethnographies, video studies, and various other scrutinies.

Despite significant differences among researchers from various disciplinary and methodological commitments about the goals of schooling; the nature of learning; the relationship between school and society; and the influence of various factors in classroom life and student achievement, most researchers agree that: (a) classrooms are complex spaces and that student achievement, however it is defined, is influenced by a wide range of factors (Cuban, 2013); (b) policy efforts are generally premised on simplistic understandings of classrooms and do not take into account the complexity, the contingency and contextual nature of classrooms (Fullan, 2011); (c) policies have, in general, failed to influence meaningful change, and classrooms have proven to be remarkably intractable and resistant to change (Payne, 2008); and (d) successful school reforms require the engagement of the school community – especially teachers – working collaboratively to diagnose issues, develop and implement solutions, and evaluate the progress of their efforts (Horowitz, 2005).

Education 2030 takes an even more comprehensive approach, rightly looking beyond the classroom to a range of other societal factors, from security to health and nutrition as impacting education outcomes. Despite this, there is a tendency, in contemporary accountability practices, to focus on defining the role of each actor as if that actor acted independently, identifying key performance indicators for their role and holding individuals accountable against set criteria. Such clarity and order at individual levels are viewed as key for successful accountability. Nowhere is this more evident than in the host of efforts devoted to measuring and improving 'teacher quality', which is conceptualised as inherent in the nature of individual teachers. The isolation of the effects of specific and individual factors is aided by statistical analysis that aims to 'control for' the effects of poverty, ethnicity, gender and so on. Two major issues can be raised with such analyses – one, where these factors interact in complex ways, 'isolating' them statistically distorts the picture; and two, even if these factors are isolated for the purposes of calculation, they present themselves in all their inter-related complexity in real classrooms. A better understanding of the networked nature of education practices and the complex interactions that produce education outcomes is needed. The unit of analysis of student outcomes should encompass at least the most major components that influence outcomes, taken as a complex collective, rather than parsed as individual factors whose influence can be isolated.

'Formative accountability' should be extended all along the chain upstream, instead of only at the classroom/school/outcome level. If strategies have been advocated for funding, or developing new teacher preparation programs, then the extent to which these strategies are being followed and are in fact useful also need to be studied - not just the end-points or outcomes. The unit of analysis for accountability needs to be conceptualised in a much more lively and dynamic way, and accountability understood as an evolving process of learning.

6.3 Balance external and internal accountability

Even where collectives take responsibility for improving education outcomes and all is going extremely well, some level of external accountability is desirable to protect citizens and ensure that communities are not disadvantaged by inefficient or corrupt practices. External accountability measures can take the form of periodic, sample-based assessments that are focused on assessing *systems* rather than individuals.

In countries like Australia, full-cohort comparative assessments have been instituted. NAPLAN surveys student learning at four points within the years of compulsory schooling. Because they are full-cohort studies, even though they are standardised and as such limited in the data they can generate on individual students, and even though classroom teachers would know a lot more about individual students than the tests could reveal, teachers are encouraged to trust NAPLAN as providing useful and impartial data to inform teaching and learning. At the same time, classroom-based, full cohort surveys link teachers directly to the performance of their classes, leading to unhealthy comparisons, impoverished and unfair evaluations, and the possibility of cheating to ensure that their class performs well. Large-scale assessments are limited in their ability to assess individual students, schools and teachers in a useful or effective way, so expensive, full-cohort data are not justified from the perspective of producing information about students either.

System level testing should not only be sample based, but infrequent. Not that much appreciable change occurs from one year to the next – so surveys conducted every other year might suffice. This would reduce costs not only in terms of money, but also in terms of misuse of data and the distraction and poor policy decisions that may emanate from frequent, full-cohort assessments.

6.4 Develop tolerance for uncertainty

Anxiety to impose accountability regimes that are tough and uncompromising, such as the *Framework* appears to reflect, invite a focus on implementation, consequences and follow-through. They assume that control based on predictability, precise measurement and correct training are essential for enforcing reform and improvement. However, as I have argued elsewhere, uncertainty can be an asset for both policy making and the development of better practices (Gorur, 2014b; Gorur & Koyama, 2013). In Australia, the introduction of NAPLAN and the like-school comparisons based on NAPLAN were meant to narrow the scope of argument with regard to classroom and school practices, and to produce definitive 'best practices' that could then be spread to all schools in Australia. But such certainty closes down spaces of exploration and prevents the opening up of new scenarios.
A tolerance for uncertainty opens up the possibility of a range of research practices that could point to new understandings of how different factors mitigate against change. Such a tolerance can also improve practices of statistical analysis that are currently burdened with producing definitive information that is beyond the scope of their epistemic possibility.

6.5 Encourage 'informed publics' to inform policy and practice

A role is envisaged for citizens, NGOs and other organisations in Education 2030's vision of participatory accountability. Currently such accountability is visualised in punitive terms – as exposing poorly managed schools and inefficient teachers and principals. A collective understanding of the nature of education reform would envisage parents and the general public also as implicated and involved in the process of change.

Public accountability needs to be visualised in more participatory ways than merely finger-pointing or holding education institutions to account. Callon, Lascoumes and Barthe (2009) provide many instances of how democratic forums came to be developed, where citizens extended the policy imagination by producing new scenarios and raising new questions and even legitimising new ways of thinking which might be have earlier been dismissed as irrational or uninformed. In my forthcoming work, I am extending these understandings based on empirical work in Australia, and exploring how numbers might be 'refused', in part through the work of 'informed

publics' (Gorur, forthcoming). Such participation would hold accountability practices themselves accountable, and question the limits of the accountability approaches being employed.

Education 2030 could work with local NGOs and other groups to mobilise and strengthen these kinds of participatory, bottom-up approaches to accountability.

7. Focus for further research

The route to participatory, collective and sustained accountability requires some further research to understand how reform processes occur and how various actors interact to influence change. Some areas for further research are outlined below.

The effectiveness and policy uptake of assessment models

With the alarming statistics of 250 million children globally still being unable to read or count, among whom half have been in school for four years or more, there is universal agreement that improving both the *provision* and *quality* of education in the Global South is of critical importance, and that for such improvement to occur, both provision and quality must be monitored through systematic and regular *assessments*. The achievement of measurable learning outcomes has become the main priority for education development in the post-2015 policy discourse (Barrett, 2011). However, while there is agreement that some instrument for informing policy and monitoring progress is required, there is widespread *disagreement* about the nature, scale, frequency and even purposes of such assessments. What kinds of indicators should be used? At what ages should the major assessments occur? Are national assessments more useful, or is there a case for regional or international comparisons? Should the assessments be done in school or as household surveys (since many students are not enrolled in schools in these countries)? Should the assessments be of basic skills, or will that set the bar too low and allow governments to ignore quality? Should all children be assessed or only a sample? At what level of disaggregation will data become useful to policy makers and practitioners? Decisions on these issues can be highly consequential, as they can dramatically change understandings of the policy issues and thus influence the deployment of funds and policy attention, with very real consequences for millions of children. A wrong decision could see inequity increase and outcomes worsen for the most vulnerable and marginalised children.

There are already a number of regional comparative assessments of education in many parts of the world, developed and deployed at considerable expense, including Southern African Consortium for Monitoring Educational Quality (SACMEQ); the Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (PASEC); El Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE); and the Pacific Island Literacy Levels (PILL) Survey. In fact, there are now 136 national large-scale assessments around the world, in addition to 171 high stakes tests (Cheng & Omoeva, 2014). There has been little research on the extent to which these expensive assessment exercises are able to inform policy or serve as instruments of government accountability. Whilst donors of education aid are keen to invest in accountability measures to ensure that their aid investment produces the desired results, creating yet more assessments, without an understanding of how these assessments operate, what kinds of information they are able to generate and what policy lessons they are able to mobilise would not only be pointless, but a travesty that would lead to continued poor outcomes for another generation of children.

Research that investigates different assessment programs to examine how different approaches and methodologies impact the twin aims of informing governments as well as enabling accountability is urgently needed. Studies that examine diverse approaches – for example, PISA for Development, based on robust statistics, faith in mathematics and quantitative expertise, and Pratham's Annual Status of Education Report (ASER), which is a citizen-led household survey using volunteers and a very rough assessment instrument, would

generate different pictures of 'collectives' that emerge through such exercises. The idea is not just to find 'best practices' in assessment; rather, it is to understand what factors should be considered when assessing the effectiveness and influence of different assessment regimes.

Mundane objects and accountability

Woolgar and Neyland (Woolgar & Neyland, 2013) demonstrate how 'mundane' objects such as colour coded garbage bins to encourage recycling can regulate citizens' practices and bring about significant change. In education contexts, such objects may include certain forms of collective interactions facilitated by meeting formats or certain practices of data generation that are part of the taken-for-granted processes of governance. Stephen Ball has studied how certain 'policy objects' such as 'school inspections' can regulate various practices of schooling. I have demonstrated (Gorur, 2015b) how such policy objects as NAPLAN and the public website My School which publishes a range of data about schools in comparative formats can act as 'interesting objects' that gather towards themselves actors who might otherwise not have engaged with the issue of school reform; and how they mediate new relations between a range of actors. These theoretically innovative ideas may be used not only to understand current phenomena, but to actively create desired scenarios.

Developing complex accounts of 'what goes on in classrooms'

The OECD has identified student learning as 'ultimately the product of what goes on in classrooms' **(**OECD**,** 2010, p. 3). Classrooms have been subjected to qualitative and quantitative analyses, longitudinal studies, ethnographies, video studies, and various other forms of scrutiny. As a result, descriptions of classrooms abound: statistical studies have identified how particular aspects – high-stakes testing, student motivation, the teaching of phonics, formative assessment, teacher accountability schemes – influence learning. Psychologists have explained success and failure at school on the basis of individual characteristics. Sociologists have explored the relationship between school and society.

Despite so much being known about classrooms, reform efforts typically meet little success, and classrooms have remained intractable (Payne, 2008). In recent years, Australia has made significant efforts to focus on 'what goes on in classrooms', and to improve teacher quality and raise student performance. It has invested $2.5 billion under the Smarter Schools National Partnerships over the last five years. Other investments include the annual national testing of students through the NAPLAN and the development of the national curriculum and new national teacher standards. Between 2000 and 2009, real expenditure on education in Australia increased by 44 per cent (Jensen, 2012). Despite these efforts, there has been little improvement in student scores on national or international assessments. A recent Council Of Australian Governments (COAG) Reform Council report (2013) found that between 2008 and 2012, there has been no change in reading scores in Years 7 and 9 or numeracy scores in Years 3 and 9, and there was a significant *decline* in Year 7 numeracy scores. Of greater concern is that inequity has increased (Thomson et al., 2013). Whether or not one considers scores on NAPLAN to be important, it is baffling that these scores have not improved despite the focus and the investment in improving them.

The complexity of classrooms has been explored by many ethnographers. Drawing upon a mélange of methodologies during his two year ethnography, Jackson (1968) noted that the roles played by teachers and students are contingent upon 'macro' issues such as system-level factors, as well as 'micro' issues located in the classroom. He found that the 'daily grind' of classroom routines, often ignored as 'noise' in some studies, played a significant role in classroom life, carrying the 'hidden curriculum' of the classroom, and forming a normative layer that underpinned the more visible aspects of the curriculum, such as texts and tests. He recommended that the technocratic or 'engineering' view of schooling needed to be replaced with a much more complex model, because classroom life resembled 'the flight of the butterfly' rather than the 'flight of the bullet' (p. 166).

More recently, Kennedy (2005) set out to 'go as deeply as possible into classrooms, and into teachers' heads' (p. 2) to explore why teachers seem immune to reform efforts. She demonstrated that schools have become much more contested and now involved a number of competing actors:

> *[N]numerous groups try to influence teachers' local decisions. Textbook publishers, professional associations, parent associations, religious groups, and business alliances all enter the education arena. Some compete with state and local policies, some try to change them, and some reinforce them. The education policy landscape is routinely crowded with competing, often conflicting, ideas about what is the most important thing for teachers to accomplish in their classrooms. (Kennedy, 2005, pp. 3-4)*

While ethnographers have noted that classrooms are complex and idiosyncratic, Cuban (2013) contends that reform plans are typically premised on models of classrooms as *complicated* rather than *complex.* To use his lively analogy, policy makers see schooling as 'a Boeing 737 to be piloted using algorithms and flowcharts', rather than 'a complex, dynamic, and … messy multi-level system such as air traffic control with controllers adapting constantly to varying weather conditions, aircraft downtime, and daily peak arrivals/departures of flights' (pp. 114-115).

Much more research is required to describe the complexity of classrooms if we are to develop strategies to overcome the intractability of the problems that beset education. Accountability instruments can then take into account these complexities to produce useful strategies that promote effective practices. These complex inter-relations would also hold promising lessons for collective accountability.

Research on local models of successful participatory practices

There are many instances of individual success stories of turn-around schools that also include participatory and citizen-led approaches to school reform and accountability. However, their lack of scalability is seen as a major drawback – it does not need to be. Research into such collective approaches and a cataloguing of exemplary efforts may be very useful in inspiring other, locally adapted projects which do not aim to replicate models that have succeeded elsewhere, but to develop new models using these exemplars as inspiration.

8. Conclusion

The task before Education 2030 is daunting in its scope and ambition. Perhaps in response to the scale and complexity of the task before it, it has visualised strong measures to hold actors to account to ensure that its goals are met by 2030. This strong approach to accountability as regulation and control contrasts with the collective, participatory and sustained nature of education reform that is implicit in the *Framework* and the strategies outlined within. This paper proposes that Education 2030 should complement its vision of participatory, collective and sustained efforts to improve education outcomes with accountability measures that are also participatory, collective and sustained. An external, top-down approach that emphasises rules, benchmarks and compliance would be detrimental in a process that requires experimentation, involves uncertainty and requires on-going learning. More research is required to elaborate how such collective processes work and also on how different approaches to assessment work in terms of the lessons they generate for policies and practices.

References

Asdal, K. (2011). The office: The weakness of numbers and the production of non-authority. *Accounting Organizations and Society, 36*(1), 1-9. doi: DOI 10.1016/j.aos.2011.01.001

Banks, G. (2009). Evidence-based policy making: what is it, and how do we get it? In J. Wanna (Ed.), *Critical Refelctions on Australian Public Policy - Selected Essays* (pp. 107-113). Canberra: ANU E Press.

Barber, M., & Mourshed, M. (2007). How the world's best-performing school systems come out on top: McKinsey and Company.

Barrett, A., M. (2011). A Millennium Learning Goal for education post-2015: a question of outcomes or processes. *Comparative Education, 47*(1), 119-133.

Barry, A. (2002). The Anti-Political Economy. *Economy and Society, 31*(2), 268-284.

Bottani, N. (1994). The OECD International Education Indicators. *Assessment in Education: Principles, Policy and Practice, 1*(3), 333-350.

Bottani, N. (1996). OECD International Education Indicators. *International Journal of Educational Research, 25*(3), 279-288.

Bottani, N. (2008). The oil level, the engine and the car: The stakes involved in assessing the quality of education on the basis of indicators. *Éducation et Société, 18*(2), 11-23.

Bovens, M. (2006). Analysing and Assessing Public Accountability: A Conceptual Framework *European Governance Papers (EUROGOV)*.

Breakspeare, S. (2012). The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance *OECD Education Working Paper* (Vol. 71). Paris: OECD.

Buchanan, J., Prescott, A., Schuck, S., Aubusson, P., Burke, P., & Louviere, J. (2013). Teacher Retention and Attrition: Views of Early Career Teachers. *Australian Journal of Teacher Education, 38*(3).

Callon, M., Lascoumes, P., & Barthe, Y. (2009). *Acting in an uncertain world : an essay on technical democracy*. Cambridge, Mass.: MIT Press.

Cheng, X., & Omoeva, C. (2014). EPDC National Learning Assessments Mapping Project: Key Findings. Washington DC: Education Policy Development Centre.

Cuban, L. (2013). Why so many structural changes in schools and so little reform in teaching practice? *Journal of Educational Administration, 5*(12), 109-125.

Erkkila, T. (2007). Governance and Accountabiity - A Shift in Conceptualisation. *Public Administration Quarterly, 31*(1), 1-38.

Fullan, M. (2011). *Choosing the Wrong Drivers for Whole School Reform.* Paper presented at the Seminar Series.

Gorur, R. (2011a). ANT on the PISA Trail: Following the statistical pursuit of certainty. *Educational Philosophy and Theory, 43*(S1), 76-93.

Gorur, R. (2011b). Policy as Assemblage. *European Educational Research Journal, 10*(4), 611-622.

Gorur, R. (2011c). *Policy Assemblage in Education.* (PhD), The University of Melbourne, Unpublished.

Gorur, R. (2013). My School, My Market. *Discourse-Studies in the Cultural Politics of Education, 34*(2), 214-230. doi: Doi 10.1080/01596306.2013.770248

Gorur, R. (2014a). Producing calculable worlds: education at a glance. *Discourse-Studies in the Cultural Politics of Education, 36*(4), 578-595. doi: Doi 10.1080/01596306.2015.974942

Gorur, R. (2014b). Towards a Sociology of Measurement in Education Policy. *European Educational Research Journal, 13*(1), 58-72.

Gorur, R. (2015a). Assembling a Sociology of Numbers. In M. Hamilton, B. Maddox, & C. Addey (Eds.), *Literacy as Numbers – Researching the Politics and Practices of International Literacy Assessment* (pp. 1-16). London: Cambridge University Press.

Gorur, R. (2015b). The Performative Politics of NAPLAN and My School. In G. Thompson, S. Sellar, & R. LIngard (Eds.), *National Testing and its Effects: Evidence from Australia*. London: Routledge.

Gorur, R. (2015c). Seeing like PISA: A Cautionary Tale about the Performativity of International Comparisons *Working Paper #1*.

Gorur, R., & Koyama, J. P. (2013). The struggle to technicise in education policy. *Australian Educational Researcher, 40*(5), 633-648. doi: DOI 10.1007/s13384-013-0125-9

Gorur, R., & Wu, M. (2015). Seeing like PISA: A cautionary tale about the performativity of international assessments. *Discourse-Studies in the Cultural Politics of Education, 35*(5), 647-664.

Hattie, J. (2003). *Teachers Make a Difference: What is the Evidence*. Paper presented at the ACER Annual Conference - Building Teacher Quality, Auckland.

Hooge, E., Burns, T., & Wilkoszewski, H. (2012). Looking Beyond Numbers: Stakeholders and Multiple School Accountability *Education Working Papers* (Vol. 85).

Horowitz, J. (2005). *Inside High School Reform: Making the Changes that Matter*. San Francisco, CA: WestEd.
Jackson, P. W. (1968). *Life in Classrooms*. New York: Rinehart and Winston.

Jacob, B. A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in Chicago Public Schols. *Journal of Public Economics, 89*(5-6), 761-796.

Jensen, B. (2012). Catching up: Learning from the best school systems in East Asia. Melbourne: Grattan Institute.

Kearns, K. (1996). *Managing for Accountability: Preserving the Public Trust in Public and Nonprofit Organizations*. San Francisco: Jossey-Bass.

Kennedy, M. (2005). *Inside Teaching - How Classroom Life Undermines Reform*. Cambridge, MA, London, England: Harvard University Press.

Koliba, C. J., Mills, R. M., & Zia, A. (2011). Accountability in Governance Networks: An Assessment of Public, Private, and Nonprofit Emergency Management Practices Following Hurricane Katrina. *Public Administration Review, 71*(2), 210-220. doi: 10.1111/j.1540-6210.2011.02332.x

Latour, B. (2004). Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern. *Critical Inquiry, 30*(Winter), 225-248.

Lingard, R., Martino, W., Rezai-Rashti, G., & Sellar, S. (2016). *Globalising Educational Accountabilities*. New York: Routledge.

Meyer, H.-D., & Benavot, A. (2013). *PISA, Power and Policy*. Oxford, UK: Symposium Books.

Mulgan, G. (2005). Government, Knowledge and the Business of Policy Making: The Potential and Limits of Evidence-based Policy. *Evidence and Policy, 1*(2), 215-226.

Mulgan, R. (2000). 'Accountability': An Ever-Expanding Concept? *Public Administration, 78*(3), 555-573. doi: 10.1111/1467-9299.00218

Nichols, S. L., & Berliner, D. C. (2007). *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Cambridge: Harvard Education Press.

Normanton, E. L. (1966). *The Accountability and Audit of Governments*. Manchester: Manchester University Press.

Norris, J. J. (2014). Rethiking accountability in new governance. *Innovation: The European Journal of Social Science Research, 27*(3), 199-219.

OECD. (2010). PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Vol. IV). Paris: OECD.

Parsons, W. (2002). From Muddling Through to Muddling Up - Evidence Based Policy Making and the Modernisation of British Government. *Public Policy and Administration, 17*(3), 43-60. doi: 10.1177/095207670201700304

Payne, C. M. (2008). *So Much Reform, So Little Change: The Persistence of Failure in Urban Schools*. Cambridge, MA: Harvard Education Press.

Porter, T. M. (1995). *Trust in numbers : the pursuit of objectivity in science and public life*. Princeton, N.J.: Princeton University Press.

Power, M. (1996). The Audit Explosion. London: Demos.

Power, M. (1997). *The audit society : rituals of verification*. Oxford ; New York: Oxford University Press.

Ranson, S. (2003). Public accountability in the age of neo-liberal governance. *Harvard Educational Review,*

72(3).

Romzek, B. S., & Dubnick, M. J. (1987). Accountability in the Public Sector: Lessons from the Challenger Tragedy. *Public Administration Review, 47*(3), 227-238.

Sauvageot, C. (2008). A Tool for International Comparison: The International Standard Classification of Education (ISCED). In A. Charras & C. Sauvageot (Eds.), *Education and Formations: International Comparisons*: Ministry of Higher Education and Research Evaluation, Forecasting and Performance Directorate.

Scott, J. C. (1998). *Seeing like a state : how certain schemes to improve the human condition have failed*. New Haven: Yale University Press.

Smith, A. (2015, 5 August). NAPLAN 2015: Education chief warns students are not improving. *The Sydney Morning Herald*.

Strathern, M. (2003). New Accountabilities - Anthropological studies in audit, ethics and the academy. In M. Strathern (Ed.), *Anthropological studies in accountability, ethics and the academy*. Florence: Taylor and Francis.

Thomson, S., De Bortoli, L., & Buckley, S. (2013). PISA 2012: How Australia Measures Up. Camberwell, Victoria: Australian Council for Educational Research Ltd.

UNESCO. (2013). Concept note on the Post-2015 education agenda. Paris: UNESCO.

UNESCO. (2015a). Education 2030 Incheon Declaration and Framework for Action. Paris: UNESCO.

UNESCO. (2015b). EFA Global Monitoring Report 2015 - Education for All 2000-2015: Achievements and Challenges. Paris: EFA Global Monitoring Report.

UNESCO. (2015c). Thematic Indicators to Monitor the Education 2030 Agenda - Thematic Advisory Group Proposal: UNESCO.

Woolgar, S., & Neyland, D. (2013). *Mundane Governance: Ontology and Accountability*. Oxford: Oxford University Press.

[i] https://bctf.ca/IssuesInEducation.aspx?id=5724
[ii] http://www.edweek.org/ew/toc/2016/01/07/index.html?cmp=eml-sb-sr-qc16-20160107Report
[iii] https://www.det.nsw.edu.au/proflearn/docs/pdf/qt_hattie.pdf