UKFIET and the Overseas Development Institute (ODI)
Roundtable on Post-2015 Goals, Targets and Indicators for Education
1 May 2014, 2-5 pm at the ODI

INDICATORS OF QUALITY IN TEACHING AND LEARNING: WHAT REALLY MATTERS?

Robin Alexander

This brief contribution to the ODI/UKFIET roundtable draws on but also goes beyond a longer keynote given at the Oslo launch of GMR 2014 on 3 February 2014: http://www.robinalexander.org.uk/wp-content/uploads/2014/02/Alexander-Oslo-GMR-2014B.pdf . The Oslo paper includes most of the references alluded to here.

The various and sometimes competing post-2015 draft targets and indicators for education make 'quality' a mantra that begs questions every time it is invoked, which is rather too frequently: 'quality early childhood care and education', 'quality basic education', 'quality upper secondary and tertiary education' ... So, to recast slightly what I asked in my Oslo commentary on GMR 2014:

- Will the post-2015 account of educational quality attend to what in learning and teaching really matters?
- Will the classroom processes that for children and young people are truly transformative be adequately captured in the post-2015 goals, targets, indicators and measures used, and in the evidence on which monitoring and policy are based?

Because I believe that a wholly understandable emphasis on input and outcome in successive GMRs has been associated with conspicuous neglect of school and classroom processes, it's on process that I've tended to concentrate in my two GMR commentaries, one for DfID in 2008, the other for the Oslo launch of GMR 2014 this February. For process *matters.* Certainly, as in one of the draft indicative national targets, children should access and complete their primary education, but what happens during those five or six intervening years which psychologists and neuroscientists tell us are, after the early years, the period of most rapid and significant human development and learning? Of course children should leave primary education 'able to read, write and count well enough to meet minimum learning standards' but how do they reach those standards? And how does this laudable target square with that of achieving 'a broad range of learning outcomes', which presumably means rather more than the 3Rs? And what do we say about the apparent dissonance between one statement's exclusive emphasis on functional numeracy and literacy while elsewhere schools, echoing GMR 2014, are also charged with educating for the much more grander goals of gender equity, sustainability and global citizenship.

Further, process – teaching and learning - can capitulate to socio-economic pessimism, or it can challenge and even defeat it. Researchers agree, give or take a few percentage points, that school effects may account for only 30% of a student's attainment, and this reminds us that what happens outside school is at least as influential for the child's learning and life chances as what happens within it. Yet we know that the odds in the context of disadvantage and deprivation can be massively shortened by teaching which is outstanding rather than merely routine. 'Good teaching makes a difference,' we noted in the Cambridge Primary Review, 'but excellent teaching transforms lives.'

For all these reasons, a careful engagement with educational process is as essential to education post 2015 as it ought to have been up to now but hasn't. Conversely, a skewed or inadequate

conceptualisation of educational process will distort everything else, from the targets set to the indicators formulated, the measures used and the policy and funding decisions which determine and follow these targets, indicators and measures. So, for example, if we don't know wherein the quality of *teaching* resides, we are not in a position to construct high quality teacher training as rightly commended in GMR 2014.

It is of course a question of balance. Process matters, as I've said, but so does input, especially the quality of teachers, and we'll get nowhere without a valid vocabulary of learning outcomes. But the proper explication of both of these depends in part on the explication of process.

Here, then, are my main concerns about the story so far:

- In the input-outcome model, process, and the contribution of process to outcomes, tend to be inferred rather than examined.

- EFA quality indicators appear to have been chosen on the basis of availability and measurability rather than validity. Since relatively little in education is amenable to measurement, this almost certainly means that vital inputs, processes and outcomes have been ignored not just in tracking the progress of EFA but also in shaping it. (This is not an objection to measurement as such. For those dealing globally with very large populations as the GMRs 2002-14 have done, quantitative data are both the most useful and most appropriate means of tracking and representing trends, and to reduce the problem of accessing process to that tired old quantitative vs qualitative paradigm war is unhelpful. Yet every method has its limits, and methodological exclusivity can be dangerous).

- Proxies for quality are inevitable but they may be used to excess and some have a questionable relation to that for which they purport to stand. What, for example, does that unfortunate 'survival rate to grade 5' really measure apart from, well, survival to grade 5? The same might be asked about the less desperate term 'completion'. And do 'survival' and 'completion' reliably signal that what precedes them is educationally worthwhile?

- The word 'quality' in EFA literature slides almost nonchalantly between the adjectival and nominal, between the descriptive and normative, between statements of provable fact and assertions of unsubstantiated value which rapidly dissolve into slogans. 'Activity methods' and 'student-centred teaching' are typical examples. And what is 'quality education' other than a rhetorical kite-mark? (The linguistic problem here is that 'quality' started life as a noun and as such it needs to be modified adjectively to form the basis of an indicator or measure, but in common with an increasing tendency fostered by transatlantic corporate-speak 'quality' is often treated as an adjective and at that point becomes meaningless).

- In similar vein, the terms 'indicator' and 'measure' are too frequently confused or elided. Here I realize I may be in a minority, for an indicator was defined at this session as 'a precise metric from identified databases that assesses if a target is being reached', and nobody disagreed. But that definition is stipulative, not absolute, and I prefer to exploit the larger possibilities allowed by our language to argue that if we have the luxury of two terms they should not squandered by treating them as synonymous. Measures measure; indicators indicate. They do different jobs. Dark clouds indicate rain but they don't measure rainfall. A noisy classroom may indicate impending behavioural chaos, engrossed excitement or neither, and while noise can be measured by decibels, decibel levels don't infallibly measure levels of learning. For reasons we know and understand, what can be measured is privileged over what cannot, but that's a dubious elevation if what is measured has limited indicative power and what is important is

ignored. Some indicators can be translated into measures, some cannot, but let's talk about the full spectrum of what *needs* to be indicated before we start talking about measures.

- The most conspicuous absentee from this entire discourse is pedagogy, or the craft, art or science of teaching and the evidence that sustains it.

Why are we in this fix?

- The analysis offered by the US National Research Council (NRC), which arose from a consideration of the use of international evidence to inform US national policy, applies equally to the case of EFA and the agencies that inform it. That is, policy makers tend to be interested only 'Type 1' data that is macro, systemic and quantitative; or in simple and often simplistic cause-effect 'Type 2' extrapolations from these data of the kind produced by McKinsey. Such data arise from studies that *infer* process causes from input and output data but rarely investigate them as fully as they require. Meanwhile the vast array of studies that do investigate process in depth, classified by NRC as 'Type 3', are ignored on the grounds that they are too small, too 'soft' or both. Which is a pity, say NRC, for these are the studies which access causality in practical school and classroom terms rather than merely chart national and international trends. Thus it is that in England Secretary of State Gove claims that the way to raise standards is to copy the *prescribed* maths curriculum of Singapore, ignoring Type 3 process research from Singapore itself that demonstrates that it's the *enacted* curriculum – what teachers do with what is prescribed - that makes the real difference.

- Thus it is, too, that in a GMR entitled *Teaching and Learning: quality for all*, a mere six per cent or so of the cited references appear to deal with teaching and learning as they happen, a larger proportion deal with teacher supply, training and retention but the majority are macro-level national or cross-national studies of education policies, programmes, strategies, governance, funding and outcomes.

- These – the inadequate take on quality, the neglect of process, the elision of indicators and measures, the neglect of a vast swathe of relevant literature and evidence - are the symptoms. The main cause, we might infer, is the agenda dictated by policy and finance. Big bucks need big data. Another might be the skewed profile of the expertise with which national and international agencies are as a consequence stocked or on which they choose to draw. Even in relation to exercises purportedly dealing with classroom processes, the economists and policy analysts - essential people both - predominate while classroom researchers are conspicuously absent. The perspective is top-down and system-oriented. But teaching and learning are complex processes that must be engaged with directly, intensively *and locally* to be understood.

What is to be done?

- First, education for the period post-2015 needs a radical and properly informed debate about indicators and measures in relation to the black box, or black hole, of teaching and learning, for classrooms are the true front line in the quest for quality in education. The proper sequence, surely, is not to make do with the odd measure that happens to have featured in a number of school effectiveness studies – student time on task is a familiar but notoriously unreliable example – but to start with a rounded account of the educational process and the purposes it serves, then range comprehensively and eclectically across the full spectrum of relevant research and extrapolate what the evidence shows can safely be regarded as key indicators of quality, and only then proceed to the question of how those indicators that have been shown to have pre-eminent influence on the quality and outcomes of learning can be translated into measures. In all cases, both indicators and measures should resonate clearly with goals. It's all too common for

education goals to grandly espouse x and then signal through what is tested that the true priorities are rather different.

- Second, where an indicator has empirical provenance but cannot readily be translated into a simple measure, other ways should be found to keep it in the frame. Under no circumstances should an indicator that peer-reviewed research has shown to be critical to effective teaching and standards of learning be dropped at this stage merely because it can't be quantified. We have to find other ways of handling it. We need a more creative and less doctrinal approach to the whole question of indicators and measures, exploiting, as I suggest above, the methodological possibilities that the vocabulary encourages. In any case, the objectivity of quantitative measures is often overstated, while there are established procedures for assuring inter-judge reliability in the use of so-called subjective assessments. Again, I warn against paradigm wars.

- Third, to cover the evidence as it needs to be covered, teams working on the defining and monitoring of quality in education post 2015 should become more relevantly multi-disciplinary than, in the EFA context, they appear to have been thus far.

- Fourth, let's accept that although much Type 3 evidence comes from and relates to high-income jurisdictions and systems, the recent DfID review of research on pedagogy, curriculum and teacher education in developing countries shows that there's now a fair Type 3 corpus from middle and low income contexts too. One of the post-2015 tasks, I suggest, should be to expand that corpus and make it as reliable as possible. In any case, there's sufficient evidence from cross-national studies of teaching and learning, and from Hattie-style meta-analyses of classroom research, to show that there are universals to which in any event we should attend; for example, teacher professional content knowledge (not the same as subject knowledge), the character and degree of cognitive challenge afforded by teacher-student interaction, and the quality of the information conveyed in teacher-student and student-teacher feedback.

- Fifth, having identified which processes matter most and having nominated them as essential indicators, we will find from Type 3 research that some of them are more amenable to measurement than may be thought. That goes especially for the teacher-student interaction that lies at the heart of teaching.

- Sixth, in light of all this, we need therefore need to explore targets and indicators for both learning and teaching. Learning needs a process indicator as well as an outcome one, and on the basis of what we know about the crucial conditions for learning, we might try *student engagement*. Similarly if teaching has to be reduced to just one indicator, on the basis of what we know about the characteristics of effective teaching from both Hattie's meta-analysis of studies in high income countries and the recent DfID literature review, we might try *reciprocity* in teacher-student interaction. As it happens, both engagement and reciprocity are amenable to measurement, so mine isn't a completely hopeless cause.

- Finally, here's the double and troubling dilemma. I and others want teaching and learning as they happen and are experienced to gain the prominence in education after 2015 that they deserve and urgently need if we are to make progress, and because the modality highlights targets, indicators and metrics, that means that we need to start by exploring how far what matters in teaching and learning is amenable to this treatment. But on the basis of the reductionism we've witnessed thus far, the prospect of a single global measure of the quality of teaching applied across all cultural and pedagogical contexts is nothing if not deeply alarming. Quite apart from the totalitarian resonance of such an idea (what one Cambridge Primary Review witness called a 'state theory of learning'), or the possibility - nay, probability – that the measure would be plain

wrong, teaching is a quintessentially local activity; and I say this having acknowledged above that there are empirically demonstrable universals in teaching and learning that apply across cultures and systems.

- This also means that while generalized process quality targets may be volunteered on the basis of what the evidence tells us makes a difference, it is only at classroom level that they can be feasibly monitored. The trick will be to give process and the quality of process the prominence they deserve without allowing the resulting indicator(s) to tyrannise and debase what they purport to advance and improve; and to find a way to add this essential local dimension to development and monitoring processes that are no less essentially about global and national development.

Correspondence: rja40@cam.ac.uk