Draft Only

Assessment, Public Accountability and Trust

Onora O'Neill

Assessment is meant to provide evidence of quality of performance — typically the evidence is based on judging some carefully selected sample of that performance, or something supposedly systematically related to performance. So far, so seemingly simple. But once we start thinking about the reasons *why* we assess in some ways rather than others, and *how* we use the results of assessment, complexities and even difficulties multiply. Some particularly awkward difficulties arise when the evidence provided by a system of assessment is used as a basis for holding others to account. For present purposes I shall set aside purely 'internal' uses of assessment, such as its diagnostic use, or its formative use to help the learning of those assessed, and concentrate on some pitfalls that arise when the results of assessment are used as a basis for accountability.

The information provided by educational assessment can be used for many sorts of 'external' purposes. Assessments of pupil performance can provide information that helps pupils, their parents or prospective employers to make educational and employment **choices**. It can provide a basis for **selecting** individuals for specific sorts of further or higher education, or for certain jobs, and for **qualifying** individuals to undertake specific vocational or professional activities. Those who are not selected, or not certified as qualified, are then **excluded** from those activities. Assessments of pupil performance can also be used to judge not the pupils, but those who have

taught or prepared them—or failed to do so—and to **reward** and **punish** teachers and schools accordingly. And they can also be used to **support policy arguments** for more, or better, or different, sorts of education or training, or compared with the results of educational assessment in other societies to show that we are doing enough to 'build a knowledge economy' or to survive the rigours of international competition – or, alternatively that we are not.

Assessment inevitably becomes a hot issue when it is used for external purposes of these sorts. When a university offers a place to study medicine to a pupil with a certain level of exam achievement, but not to another with lesser exam achievements, life chances are at stake. When parents use school league tables as a basis for choosing a school, or for deciding to take on the expense of independent schooling, the stakes are again very high. And when poorly performing schools are faced with special measures—or closed—or not closed, life chances for pupils and staff may once again be at stake.

One of the most common external uses of the results of pupil assessment is to hold pupils, teachers and institutions to account for the quality of their performance. Accountability is *second-order*: it uses first order judgements of performance—in this case assessments, often consisting of numerical scores that are taken to offer measures or indicators of performance —as a basis for making second-order judgements that hold those who have performed –or not performed—to account. Pupils may be offered or refused opportunities for further study or to enter certain lines of employment on the basis of this these second-order judgements; those who

prepared them—teachers and schools—may be rewarded on the basis of their pupils' good performance and penalised if they perform poorly. In itself a favourable assessment is just that: but once linked to a system of accountability it may entitle those assessed to further, valuable forms of education, opportunity or employment. In itself an unfavourable assessment is just that: but once linked to a system of accountability it may become a basis for exclusion from valuable forms of education, training, employment or promotion.

2 Accountabilities and Perverse Incentives.

It is vanishingly unlikely that the evidence provided by any given system of assessment will be useful for all the ways in which it may be thought important to holding those whose work contributed a given performance to account.

For example, the system of pupil assessment now used in schools in the UK is used to provide rather broad levels of information to universities, which they are to use as a basis for offering or refusing places on each course. The system generates ostensibly numerical scores, and rankings based on those scores. It focuses on the number of 'points' an *individual* pupil has been awarded at A level. At an earlier educational stage the focus is on the number of passes, particularly A-C passes, an *individual* pupil obtains at GCSE. These scores can also be used to work out the *average* number of points per pupil, or the *average* number of A-C marks per pupil, for a given

school, or region, or type of pupil. These average scores are thought of as useful information for holding teachers and schools to account.

However, by themselves these scores will not provide good information about educational attainment, for several reasons. In the first place, neither the individual nor the average scores are more than *ostensibly* numerical. The ways in which As —and still more Bs !—[1] may be obtained vary, and there is little convincing evidence that an A (or B, or C) obtained in one subject is, educationally equivalent to an A (or B, or C) obtained in another subject. If we cannot show—and have no reason to assume—a good match between point scores and educational attainment, it is hard to use these scores as a basis for university admission. There is no genuine unit of account. Still, points are counted

Second, since the choice of subject (apart from limited core requirements at GCSE level) is left to pupils, the arrays of scores achieved by different pupils are not strictly comparable. The fact that pupils are left to choose their options in the light of school provision and advice and their own preferences has profound effects. The ansece of a unit of acount might not matter if those applying for university admission had studied a common curriculum, since a rank ordering would then be enough to support discriminating judgement of performance for what is, after all, a positional good. However, work in the final two years of schooling in the UK is not defined by any common curriculum, and, even if it were, the simplified scores that are communicated to universities would not permit fine discrimination. So at present in the UK the assessment of sixth form pupils

[1] At GCSE a B may be obtained in mathematics by taking a less demanding exam—cf. paper by Alison Wolf.

offers a doubly unsatisfactory basis for highly selective university admissions, and so for holding pupils (or their schools) to account.

Universities that admit selectively try to compensate for the reality that assessment in terms of points cannot discriminate at the top of the range, so cannot offer adequate evidence for selective universities to choose whom they should admit. Some universities now call the arithmetic fiction that assigns the same weight to an A in all subjects into question. They may discount A levels in General Studies, or lay extra weight on subjects that require accuracy—German or Physics, for example. Some take particularly serious account of GCSE marks, which are obtained before pupils are allowed to drop too many educationally fundamental subjects, so that there is a larger element of common curriculum and a better basis for ranking students. Some use aptitude tests to augment the incomplete basis for judgement that A level point scores offer them. (Some employers make parallel moves to check literacy and numeracy skills, which they fear GCSE evidence will not have measured well.)

Moreover, any attempt to use A level point scores or numbers of A-C marks at GCSE achieved on an option-based curriculum as a basis for judging pupils who wish to go to university creates a second range of difficulties. Where pupils have a choice of options there will be incentives for them—and for parents and schools—to gravitate towards subjects where good grades are easiest to come by. Precisely because point scores at A level, or the number of A-Cs GCSE, matter for pupils and institutions—because these are elements of assessment form the basis for holding them to account — there is a strong incentive to choose subjects where higher numbers of

points are expected more confidently, and to avoid subjects where they are thought to be harder to get. The effects can be clearly seen in the decline in GCSE entries in modern languages (reckoned to be more difficult than some other subjects) in maintained schools once the subject was made optional at GCSE (by a remarkably untimely piece of educational vandalism inflicted while Estelle Morris was Secretary of State).

However, the fact that the current system provides incentives for pupils and institutions to gravitate to subjects where adequate or good grades, hence points, are perceived as more readily available, shows that second order uses of the outcomes of assessment for purposes of accountability can have perverse, certainly adverse educational implications. Similarly, where HE institutions and employers adjust recruitment to compensate for the fact many pupils self-select out of subjects which are valuable for employment or university education, even if they might have done well and gained educationally, the incentives created by using assessment outcomes for certain types of accountability have unwanted effects on central educational objectives. However, what starts as a perverse effect—the displacement of pupils into less educationally desirable courses—may in the end be corrected (perhaps we should say corrected downwards) by subsequent events. For example, once the numbers in less favoured subjects have suffered a bit of attrition, provision in many schools will be cut back, and it can then become wholly rational to avoid those subjects. In effect, the accountability tail wags the educational dog.

2. Accountability cannot Supersede Trust

But what is the alternative? Surely we cannot revert from a culture of accountability based on objective assessment of performance, to one in which we simply trust schools to ensure that pupils are coming along well, universities to admit the right students, employers to make sound appointments and policy makers to make sensible educational decisions? During the last 25 years accountability has been widely seen as a successor to trust and is now deeply entrenched in nearly all aspects of educational and professional life—and far beyond. The regulatory revolution that has transformed British life, and in particular the formerly nationalised industries and the public sector, has long since dispelled any culture of trust.

Nevertheless, the assumption that accountability is an **alternative** to relations of trust is, I believe, mistaken both in and beyond the education. The mistake has I believe been based on widespread but unconvincing assumptions about the nature of trust.

The most common misleading assumption about trust sees it as a sort of cultural glue that provides 'social capital' in high trust societies, which is missing or damaged in low-trust societies. [2] On this view, trust once squandered is hard to restore, and there is little that individuals can do if

[2] Francis Fukuyama Trust: The Social Virtues and the Creation of Prosperity Free Press NYC 1995. See especially Ch. 2 'The Idea of Trust'. Also Robert Putnam (1995) 'Bowling Alone: America's Declining Social Capital', *The Journal of Democracy*, 6:1, 65-78. Bowling Alone: The Collapse and Revival of American Community, Simon & Schuster, 2000.

they find themselves in a low trust society. Trust is based on social rather than individual virtues, and individuals in low trust societies cannot do much about it—except to replace trust with accountability. They are driven (to quote Francis Fukuyama) to "co-operating only under a system of formal rules and regulations, which have to be negotiated, agreed to, litigated, and enforced, sometimes by coercive means" [3] Once trust is dissipated, the *only way* in which we can support co-operation is by imposing formal systems of accountability—despite the fact that such systems impose large transaction costs on all economic—not to mention educational – activity, including damaging perverse incentives. On this view, systems of accountability are seen as *replacing* trust by supporting trustworthy performance, while eliminating dependence on unreliable relations of trust. There is no possibility of restoring trust, since the institutions of civil society on which it depended have been eroded or discredited.

This view of trust as a matter of culture or attitude has been widely adopted during the last 15 years. For example, it is presupposed by the numerous opinion polls that assume that trust is merely an *attitude* which people have or lack. Pollsters ask their respondents undifferentiated questions, for example whether they trust those holding specific roles —doctors, teachers, journalists. Such questions eliminate any basis for an intelligent judgement about where to place and where to refuse trust. Any intelligent person would normally say that they trust some but not others holding these roles, and that they trust them in some matters but not in others. However, the assumption that trust is a free floating attitude, a bonding, binding cultural glue, licenses the thought that each person will assign a single level of trust

[3] Ibid. p 27

to types of office holders, without relying on or judging any sort of evidence. In effect, respondents are asked to respond *on the assumption that all trust is blind*: any basis for differentiating cases, of the sort that we would rely on in daily and professional life, is assumed away. Those who start with this view of trust unsurprisingly cannot find very much to say in its favour. They tend to depict it as infantile and as a form of dependence, and as something that has no proper place in the public life of mature democracy,

The claim that trust is obsolete in social and professional life, hence to be rejected in favour of accountability couldn't, however, be further from the truth. We can't have *any* form accountability without *some* forms of trust. Those who recommend the increased oversight, monitoring of standards, recording of performance outcomes, sanctioning of poor performance and tighter contractual relations required by various contemporary forms of accountability haven't miraculously discovered forms of accountability that work without trust. Rather they invite us to trust both certain complex, often arcane, processes of monitoring; inspecting and controlling that are introduced in the name of accountability, and those who impose them. Trust-free accountability is a mirage. We should not be surprised that replacing trust with accountability, life world with system world, only pushes the question of where to place and where to refuse trust further back. We need to ask of any system of accountability why it should—or should not—command our trust. The various systems of accountability that use the outcomes of educational assessment are no exception. We need to ask whether and when we have reason to trust them.

This point I think throws considerable light on the fact that there are few signs of any end to the supposed crisis of trust. The remedy for this supposed crisis has aimed to supersede rather than restore trust. It has introduced massively complex systems of accountability, to which pupils, professionals and the institutions in which they work are now held. [4] However since the systems of accountability imposed are themselves of high complexity, even obscurity, they rightly do not command automatic public trust. Yet if they are not trusted they will not meet with public acceptance. In education as elsewhere, over-complex systems of accountability, and in particular systems that create perverse incentives and frustrate serious educational objectives, are often **a source rather than a remedy** for mistrust.

2. Trust in Assessment

Perhaps, however, we can be more optimistic about educational assessment, than we can about the systems of accountability that use it. After all, each of us has been at school and taken exams, and understands what educational assessment is meant to do. Systems of accountability that are based on assessment evidence are likely to be easier for most of us to understand than systems used for holding insurers or bankers or broadcasters to account. So we might hope.

However, there are also reasons for being less optimistic. Many of us do not in fact understand how exams are marked, or what those who mark them aim to measure, or why the exams do not offer better evidence for selection

[4] Michael Power The Audit Society; Michael Moran The British Regulatory State.

for university admission or for employment. Anybody who deals with university admissions will have had heart breaking correspondence with pupils—or parents—who cannot believe that what they see as amazing performance has not met with success, and who may not be able to imagine what a better performance could be. In such circumstances, some suspect that there must have been prejudice or bad process— the outcome of the process is seen as unacceptable, and so the process is not trusted.

This may seem a small and local problem arising because in the UK we have not permitted universities to have the full assessment information, with the effect that they are prevented from making distinctions near the top of the range. Hence their increasing use of aptitude tests. However it is worth considering the possible remedies closely. If the systems by which pupils are assessed and held to account inflict educational damage, should we introduce more of the same? If pupils are *already* over examined and already over-incentivised, and schools *already* marginalising some educationally valuable activities for the sake of exam assessments**,** what will be the effects of sharpening incentives yet further? Do we want to increase the pressure of examination and assessment? Will that further educationally important objectives? Would we find it acceptable to make comparative performance more public—for example, publish an order of merit at A level or at University entrance? (You may say that the most selective universities in the US do something close to this, since they can access rank-in-class information to augment SAT scores). Would we find it acceptable if the result of making numerical information available damaged 'access'? Would we find it acceptable if 'access' were then maintained by explicitly limiting meritocratic admissions? What would that do to the

incentives that assessment systems create for pupils and schools? Should we expect litigation from schools that find that excellent performance is not rewarded for social reasons? How much educational detriment will we tolerate for the sake of loading a complex system of accountability onto our assessment system?

Indeed how much educational detriment are we already tolerating for the sake of assessment? The evidence of sixth-form pupils is that a lot of time that would once have been teaching time is now diverted to assessment, and even to mock rehearsals for assessment. In the end, it seems to me that no change in assessment methods should be acceptable if its net result is educationally damaging. Assessment is **not** an end in itself. If we do not question the reasons for basing accountability for pupils, professionals, schools and university admissions on pupil performance in assessment systems, we may forget that the real objective was educational.

3. Trust in Accountability

Our focus should perhaps be less on assessment than on the conceptions of accountability for which its results are used. Although many of us think that we understand what school assessment is for—we generally think that it is for educational purposes and on occasion for selection purposes—we may be less sure that we can judge the forms of accountability based on assessment of pupil performance that are in current use. I had a startling illustration of this point after giving the Reith Lectures, when I was asked by a local journalist why I was sceptical about the school league tables. I

asked her what she thought their merit was, and was told that without them it would be impossible to be sure that one's children were at a better school than other children. It is easy to laugh at this reply. But I think she may have got matters right, even if her way of putting it was not politically correct. League tables are easily understood *because* they offer a ranking—what they offer is (at most) a comparative judgement of the merits of different schools. They do not aim to reveal which are good and which are poor schools, let alone for which children.

The simplified types of information about school performance that are derived ultimately from assessment records, and provided for public consumption, achieve this limited aim. They may not provide much more. Ofsted reports are of course another matter, although perhaps not as informative as they might be. The types of information that are perused by professionals and examination boards are yet another matter. So it seems to me not at all surprising if those who are invited to trust on the basis of information that they are themselves in no position to assess are reluctant to place their trust in certain exams. Those who are invited to trust these systems of accountability are rarely aware of their structure; nor do they know much about the technicalities of assessment. Pupils, parents, teachers, schools and the general public know little about the statistical issues that make one or another aspect of the assessment more or less reliable; may have no sense of any difference between reliability and validity; and may not know much about the systems of assessment in use. They may be well ware of the practicalities of avoiding subjects in which it is harder to get a good mark—but not of the systematically perverse educational incentives that these practicalities signal. It is true that a great deal of information is

made available to them in the name of transparency—but transparency or disclosure is a far cry from adequate communication, and may not offer much to those who have little time or expertise to help them place or refuse trust. Yet if they cannot make these judgements for themselves, they will be in a poor position to judge whether to place or refuse trust in the relevant systems of assessment, or in the forms of accountability that are built on those systems of assessment.

In making this claims I am not suggesting that current assessment practices are untrustworthy, or that the forms of accountability that build on them are all untrustworthy. Both of them may be entirely trustworthy. But they are manifestly too complex for pupils, parents, or even teachers to judge them for themselves. Hence it is unlikely that they will be trusted.

What then would we have to achieve if we are to have an exam system that it not only trustworthy, but one in which people can place and refuse trust with some confidence? As I see it, placing and refusing trust is not generally a matter of being mired in cultural glue, but of being in a position **to judge for oneself** what a pupil, a school, an educational system is achieving. It is doubtful whether the copious evidence made available about performance meets this objective.

Forms of assessment that can be trusted are urgently needed. This is not the same as saying that reliable and valid forms of assessment are urgently needed. It is a matter of ensuring that the results of assessment offer provide usable evidence to those who need to decide whether or not to trust. An intelligent form of accountability would need to offer the public, parents

and pupils evidence *which they can use* as a basis for placing or refusing trust in teachers, in exams and in schools. Such evidence would need to allow people to make informed judgements about where to place and where to refuse trust. It isn't hard to suggest some changes that could help. I offer a few:

1. Neither intelligent accountability nor trust in systems of accountability is well served by holding schools or teachers accountable for scores on 'performance indicators' that use bogus units of measurement. Serious accountability is undermined rather than supported when teachers and schools are held to account by measures like 'number of A to C grades per pupil at GCSE' or 'A-level point scores': serious professionals know that different exams make different demands, and that different pupils achieve and thrive in different ways. Numbers are useful when we have a unit of account: we can count pupils, and we can count the money in school budgets. Other things that are important for education cannot be counted, or added, or ranked because there is no genuine unit of account.

2. Bogus numbers are not just an expensive irrelevance. They also are in fact the source perverse incentives, such as incentives to invest extra effort in pupils who might get Ds, or (at their saddest) incentives to 'massage' the 'statistics' of pupil achievement rather than improve the education of those pupils. The scores are then published, but since there is no proper unit of account they provide a poor basis for the public or parents (or universities) to decide how much trust to place in a given exam result. School league tables may tell parents that some school is 'better' or 'worse' than some other (local) school *as measured by the performance indicators*:

but they don't even aim to show that school is a good school, let alone a good school for a particular child. As anybody can see, position in a league table is a *comparative* measure; it can be administratively useful, but by itself it does not guarantee quality—or of lack of quality. In educating children, as in the rest of life, we need to make serious judgements of quality rather than relative judgements of success on questionable indicators.

3. Changing the performance indicators is not likely to resolve these problems. Yet faith in performance indicators is hard to dislodge. Every time one performance indicator is shown inaccurate, or misleading, or likely to produce perverse results, some people imagine that they will devise a better one without perverse effects. Experience suggests that they are will be as wrong as those who invented the last lot of indicators. So perhaps all we can say is that systems of assessment need to dance to the tune of educational objectives—not to the tune of accountability convenience.

4. Change may be possible if we are clearer about educational aims. But if we remain fixated on the options-based approach to secondary education, that is in many ways the source of current problems, then any answer will be difficult, since it will not be easy to assign convincing weight to different subjects offered for assessment. Although there are situations in which the rational response to uncertainty is to regard all cases as having the same properties, this is not the case for examination scores either at GCSE or at A level. That assumption is undercut by prevailing assumptions that exams in different subjects make demands of varied difficulty, by the tactical choices of exam takers, teacher and schools who select options that are more likely

to produce advantageous outcomes—even if they produce educational detriment.

5. However, supposing that that the options problem could be solved, and core subjects were defined—supposing that schools taught and pupils studied to a common curriculum— then there might, over time, be ways of securing a degree of public confidence in the results of assessment. And there might be ways of building more acceptable forms of accountability on these forms of assessment. Intelligent forms of accountability combine informed judgement of what has been done, by independent bodies, whose results are intelligibly communicated. Educational assessment is far better placed to provide a basis for intelligent accountability than many areas of life. It is after all based on assessment by people who are sufficiently *informed* to judge the activity they assess (without needing to fragment in it into the bite sized chunks needed for 'accessible examinations'), sufficiently *independent* to judge it objectively, and able to report *intelligibly* to the *various* audiences to whom account is to be given.