

Assessment and fairness

Dr. Andrew Davis

Research Fellow, Durham University

School of Education

Leazes Road

Durham, DH1 1TA

a.j.davis@durham.ac.uk

We can easily imagine people amusing themselves in a field by playing with a ball so as to start various existing games, but playing many without finishing them and in between throwing the ball aimlessly into the air, chasing one another with the ball and bombarding one another for a joke and so on. And now someone says: The whole time they are playing a ball-game and following definite rules at every throw (Wittgenstein, 1953 para 83).

Introduction

Wittgenstein (1953) urges philosophers to scrutinise examples. ‘Don't think, but look!’ (para 66-7) He believes that they should embrace the fine grain of events and processes rather than forcing phenomena to fit preconceived theories. The context is his famous discussion of family resemblance. We cannot discover something common to all games. Hosts of complex similarities obtain between pairs and small groups, with no unifying thread. Similarly, our linguistic activities are richly diverse, and we should not expect them to share any one feature. In this paper I pursue my inquiries in the spirit of these injunctions. However, I also engage in argument of a more structured kind than that favoured by the later Wittgenstein.

My explorations here revolve around the following questions: What forms of educational assessment are fair to all, and in what sense or senses?

Founding this inquiry on examples is shown to be especially important when examining these issues, given the wide range of situations to which they can be applied. I begin with a number of case studies involving assessment procedures, noting how they are criticised and sometimes changed for selected groups in the light of concerns about fairness. I work with a broad notion of ‘example’, where each case includes an assessment context, together with the cultural and political circumstances in which it is embedded. I refer to the overt and tacit values behind related policies and practices.

By attaching such importance to contexts I do not mean to imply that my approach ‘leaves everything as it is’ (Wittgenstein, 1953, para 124). As far as I am concerned, both the values behind key policies and judgments on the fairness of assessment procedures may be inconsistent, confused or lack justification. Yet such opinions comprise vital data for exploration, and must be examined on their merits rather than being sidelined in the interests of predetermined theories of fairness and justice.

The later Wittgenstein eschewed systematic argument. Some commentators take this to be a radical symptom of his view of philosophy itself. Others have felt that his self-denying ordinance reflected some of his personal issues about the subject. In my view it can be perfectly legitimate to use aspects of contexts to generate argument. So here, I begin with examples but then develop some abstract lines of philosophical reasoning. The latter incorporate a critique of some of the jargon employed by empirical researchers and policy in relation to assessment. These terms include 'construct validity' and 'accommodation'. This feature of the paper is typical of much of my writing in philosophy of education. I am constantly on the lookout for unwarranted certainties, both normative and conceptual, especially where such confidence damages the interests of students and teachers.

I pursue some negative theses about the application of fairness verdicts to educational assessment. I contend that no wholly coherent and consistent approach is possible. The opposite is sometimes assumed by those bringing accusations of unfairness. Moreover, even where the fairness of a given assessment process seems to have a significant value, the latter will often need to be weighed against other concerns when taking practical decisions about assessment policy.

Disability accommodations

I begin with familiar attempts to offer fair assessments to students thought to suffer from learning disabilities. For example 'accommodations' are frequently given to 'dyslexic' students. Sometimes a rationale is supplied:

- a variation in assessment methods should be allowed for a student who can achieve the specific learning objectives but is prevented by disability or specific learning difficulty from demonstrating this through the usual assessment methods (University of Brighton, 2009).

This thinking seems to underlie most accommodations even where it is not made explicit. Now what is supposed to be 'unfair' about subjecting a dyslexic student to the same assessment process as other students? At first sight, this is a concern about the validity of the assessment process. It is believed that conventional forms of assessment fail to measure a dyslexic student's knowledge and understanding of the intended learning outcomes. Validity lapses can, apparently, be instances of unfairness.

I would argue that a necessary condition of the very coherence of this approach is that there is a clear difference between the actual possession of knowledge and understanding and the mere manifestation of it by means of successful test performance. The conventional concept of *construct validity* fits in nicely here. I understand construct validity in the usual way: a test is valid to the extent that it succeeds in measuring a relevant ‘construct’.

...a question like Are IQ tests valid for intelligence? can only be posed under the prior assumption that there does exist, in reality, an attribute that one designates when using the term intelligence; the question of validity concerns the question of whether one has succeeded in constructing a test that is sensitive to variations in that attribute (Borsboom, Mellenbergh and van Heerden, 2004, p. 1065).

So an intelligence test has construct validity if it measures how much intelligence a student has. ‘Intelligence’ itself is the construct. Constructs in the context of educational assessment are often conceptualised as ‘traits’, as unobservable but persisting aspects of mental life.

The view that validity failures linked to disabilities can be instances of unfairness sees students as possessing varying ‘amounts’ of relevant learning when tested. For the majority, conventional tests can detect the learning. However, a minority labour under certain disadvantages associated with familiar disability labels such as dyslexia. The disadvantages block the usual manifestations of learning. Hence standard assessment devices cannot sample relevant performances and so fail to ‘measure’ cognitive achievement.

Often enough, fair treatment provisions seem eminently sensible so long as we embrace the idea that manifestations of a construct can be separated from the actual possession of that construct. Consider, for instance, the use of special spell checks, tinted glasses, a PC with software designed to support those with access problems, extra time, rest breaks, scribes and someone to read the exam questions. All this is consistent with the postulation of underlying cognitive traits whose manifestations are sometimes masked by specific constraints associated with learning disabilities.

However, the posited division between construct and manifestation is put under severe strain by some proposed accommodations. Markers may be advised to discount poor spelling and grammar. This assumes that such aspects of performance are separable from the cognitive achievements the assessments are supposed to be probing. Yet ‘grammar’ issues could range from the relatively superficial to matters impinging on the very coherence of the

thinking in an answer. Coherence of thinking is unlikely to be an optional component of any kind of cognitive achievement. 'Grammar' is represented in proposed accommodations as an aspect of manifestations only. Yet on closer examination it threatens to invade the territory of the relevant constructs themselves.

So does a sharp divide between mere manifestations of knowledge comprising an underlying 'construct' and the construct itself really exist? I think not. The 'gap' between putative manifestations and the underlying construct varies from case to case. It will always be a matter of degree. I speak metaphorically here, and arguably the very notion of a 'gap' is open to criticism. Whenever we judge a piece of behaviour to be a manifestation of knowledge or ability we are interpreting it as having qualities which are bound up with mental functioning underlying it. The supposed manifestation/construct divide represents just the kind of supposed opposition between the 'inner' private mental states and the 'outer' signs of these that Wittgenstein did so much to combat. For instance, saying that a pupil answers an arithmetic problem correctly is an interpretive judgment of an agent whose thought processes cannot be stripped away from the interpretation.

Admittedly a swimming test, for instance does seem to sample the desired skills very 'directly'. Nevertheless, underlying the swimming performance *now* is, surely, a persisting complex of muscular and psychological states without which the performance would not be swimming. (Mere bodily movements might keep the body afloat, but not all bodily movements are swimming.) If the supposed divide between manifestation and underlying construct is not absolute but rather is a matter of degree, then arguably the conception of fairness as linked to construct validity is itself questionable. This is a serious threat to some standard rationales for accommodation.

Some guidelines for marking in Higher Education go even further than the typical accommodations just outlined. For example:

For a very small number of severely dyslexic students there will be a continued difficulty with presenting their work in standard academic formats, such as the continuous argument of the essay. It may be more appropriate for some dyslexic students to present their work using bullet points...

And:

The majority of students with dyslexia can.. learn to present coursework in appropriate academic formats.. However, within those forms, dyslexic students will continue to have difficulties with logical sequences of ideas and with moving smoothly from one point to another. We would ask that.. markers do not penalise work of dyslexic students (University of Greenwich, 2009).

Higher Education institutions may grant exceptions to accommodations of these kinds where it can be established ‘which learning outcomes justifiably constitute competence standards, in which case the duty to make reasonable adjustments may not apply’ (Bristol University, 2009). For instance, Greenwich implies that selected academic departments may be able to justify claims that providing ‘continuous argument’ is at the heart of their learning objectives and so establish that they should not be expected to set this element aside in the interests of treating *any* minority group ‘fairly’. The general presumption, however, is that such concessions will be the exception rather than the rule. We are to understand that normally, dealing with and presenting logical sequences of ideas can be detached from the knowledge, understanding and abilities sought by any given academic module and hence accommodations can be offered.

Is such a policy defensible? What value could learning objectives possibly have when separated from logical sequences of ideas? Even if we put these doubts to one side for a moment, is it not obvious in any case that features of some learning programmes are bound to trump considerations relating to justice and fairness? Consider a non-academic example:- There is no duty in UK law to make ‘reasonable adjustments’ to the driving test for people with impaired vision. It is perfectly legitimate to expect satisfactory test performance from all putative drivers. Fairness considerations carry no weight here. Some programmes in Higher Education, especially those with a vocational focus might well be in a similar position. If, for instance, despite strenuous efforts it turned out to be impossible to assess medical students adequately when accommodating certain kinds of disability, then, surely, accommodations should be refused. Health concerns trump those relating to fairness in the sense under consideration.

In any event, such policies depend in part on sustaining a distinction between test *modification* and test *accommodation* (Hollenbeck, 1998). Hollenbeck characterises

modification as altering the test so much that the construct it assesses is changed. He cites a reading comprehension test where students read the items. Accordingly the test is said to involve the construct of 'silent reading comprehension'. If the test were changed so that the items were read aloud for the candidates then the new construct would be 'listening comprehension'. (p. 396) I now argue that the distinction between modification and accommodation is conceptually insecure.

We are disadvantaged by our capacity to coin phrases purportedly referring to constructs. Our hazy and indefensible intuitions suggest that linguistic differences between one phrase and another mark differences between the constructs to which they refer. So we may well think, with Hollenbeck, that the construct identified by 'reading comprehension' differs from that picked out by 'listening comprehension'. But what is our criterion for deciding what counts as one construct rather than another? The ancients believed that 'the morning star' identified a different object from 'the evening star'. Astronomical science eventually remedied their error. Arguably there is nothing waiting in the wings capable of bringing off a similar feat for constructs. Some might hope that neuroscience could play this role, but Davis 2009 raises serious objections to this proposal. The arguments cannot be rehearsed here.

The widespread provision of extra time for students allegedly suffering from conditions such as dyslexia assumes the accommodation/modification distinction and that we are dealing with accommodation. What, then should be made of an argument that extra time involves *modification* rather than accommodation?

Speed of thought and response is often prioritised in ability tests developed in the West. Some other cultures take the opposite view, according to which rapid responses to a problem may well be a sign of stupidity rather than intelligence (see e.g. Sternberg 2003). When we talk in this fashion we are, on the face of it, identifying a construct that can be separated from other aspects of cognition. According to the story, certain societies value selected cognitive functioning AND speedy intellectual processing, while others prefer the cognitive functioning without the speed. If our attempts to make such references are coherent then extra time might well involve modification. Tests with extra time measure constructs *without* whatever is needed in the way of cognitive equipment for speedy thinking.

A counter-argument might run as follows. In just those cases where disabilities threaten the validity of the conventional test, modification is *not* involved after all. The candidate with the disability is held to be in possession of such and such a level of cognitive capacity. In the

conventional test the disability prevents responses from providing a valid measure. With extra time, the disability is disarmed and the test can probe the relevant cognitive capacity. Any constructs relating to *speed* of cognitive response simply do not figure. They have no chance to play any role in generating test performances. On the other hand, according to the argument under review, were candidates *without* the disability to be given extra time then mental speed traits would have a chance to influence performance. Suppose non-disabled groups in possession of equal cognitive capacities other than those to do with speed were compared. Imagine that Group A has a significant construct for speedy thinking, while Group B does not. If some of Group A are given extra time, they should perform no better than the rest of A who have the usual time. The former can think quickly and so do not benefit from having more time than their peers who also have the quick cognition construct. However, where candidates from Group B are given extra time, they might well do better than their Group B peers without the time allowance. For those lacking the speedy thinking construct, more time might well prove very helpful.

Needless to say, there are many imponderables in this argument, such as the conception of an ‘amount’ of a construct responsible for speedy cognitive functioning and whether such an amount could be ‘measured’. How could the dispute between the verdict of accommodation and of modification possibly be settled? These problems are surely *conceptual* and stem from principled difficulties with ideas of ‘construct’ and the distinction between accommodation and modification.

Let us continue to play around with the ideas of test validity and fairness for a little longer, despite our growing unease with the whole project of making sense of fairness in terms of test validity. Do other personal traits, over and above those attributable to learning disabilities, also threaten validity? For instance, some students are very anxious under test conditions. Moreover, a minority even have serious difficulties with handwriting. Cambridge University (2008) makes allowances for poor handwriting but only if this is ‘caused by a disability or medical reason’ (sic). We can only speculate on their thinking here. One account might run as follows: ‘normally’ a tacit component of *all* constructs tested at university level is the capacity to write by hand. For most examination candidates handwrite their answers. However, such a component is not regarded as central to university programmes and so can be set aside in certain circumstances.

But why should poor handwriting 'caused' by a disability be regarded as an excuse, when the same degree of incompetence lacking a cause of this kind is not? If the disability is deemed to be an excuse because the candidate *cannot help it* then surely many other non-medical cases of poor handwriting are equally beyond the control of the individuals concerned. Similar questions arise in connection with the University of Wales (2006) suggestion that students may have 'additional exam arrangements' if suffering from 'Increased anxiety as they have an underlying mental health issue'. Many would react sympathetically to this, feeling that this is 'fair' to such students because it ensures that the assessment process has the greatest chance of measuring what they know and understand. Yet, once again it is unclear why *only* states of anxiety linked to health issues trigger special treatment. A plausible case can be made that people become anxious in a wide range of situations, most of which cannot be said to involve medical conditions, and at least some of which are not under their control.

Another familiar criticism of accommodations is that they are unfair because they threaten the *reliability* of the assessment process. The special treatments offered to certain types of candidate may well not be applied consistently. If a candidate is having her question read for her, one reader may use much more appropriate expression than another (Pitoniak and Royer, 2001). Moreover, in so far as we can make sense of learning disabilities such as dyslexia, such conditions are not all or nothing. People suffer these conditions with varying degrees of severity. It is arguably 'unfair' to give a severely dyslexic student the same amount of extra time as another who is only mildly afflicted. Some dyslexic students need 'more' accommodation than others. Needless to say, it is even *more* unfair to highlight disabilities such as dyslexia if we cannot in principle be clear about what counts as suffering from such a condition, or if related existence claims lack adequate empirical evidence. I cannot pursue this issue further here.

My interim conclusion is that if accommodation policies rest on ideas of fairness linked to test validity then they are open to serious philosophical criticism. The central concept of 'construct' is flawed, and the assumed divide between the performances sampled by a test and the underlying construct supposedly measured, cannot be sustained. Moreover, even if accommodation policies based on fairness as validity could be defended, practical educational decisions must still be informed by other values too, something that defenders of the 'rights' of disability groups do not always seem to recognise.

Unfairness and the motivation of candidates

Seven years ago the National Curriculum Key Stage 2 reading tests for eleven year olds in England were described by the National Primary Heads Association as 'biased' and unfair on the following grounds: the text was factual, presented in a magazine format and offering 'bite-sized chunks' (Henry, 2001). Two years earlier, similar complaints had been made about the spider-related content, and the inclusion of cartoons. Sudden improvements in boys' scores over girls were attributed to the incorporation of material held to be more appealing to boys than girls.

First of all I will attempt to run the standard validity narrative as a way of explaining and justifying the complaint of the Primary Heads: The tests are unfair to girls because, if a boy and a girl take the test in possession of equal amounts of the relevant reading construct, the girls' lack of enthusiasm for the content means that their test performances do not fully manifest their reading capacities. In this account, lack of enthusiasm functions similarly to a learning disability.

This account is very obviously open to challenge. Why choose the explanation in terms of a deficit? Perhaps we should instead be talking about the boys' unfair advantage. Why not speak of test validity being threatened by the boys' enthusiasm for the content – they perform *too well*?

The very possibility of alternative stories here gives us a major problem. How are we supposed to choose between one verdict according to which girls have an affective deficit which undermines the capacity of the test to measure their reading abilities, and a second verdict crediting boys with an affective advantage which also damages the validity of the test? The difficulty here hints at conceptual confusions embodied in how the situation is being characterised.

Arguably there are more ways in which the original complaints could be articulated. For instance, it could be said that the test was *modified* by the presence of the boy-friendly content, the latter changing the construct concerned. After all, there is no reason why constructs should not include stable or relatively stable aspects of motivation, interest, and dispositions to find given material enjoyable. The boy-friendly test picks up on certain of these affective elements. To consider this at all we have to set aside the problems with the accommodation/modification distinction.

Such a change in construct might be good or bad, but is it obviously *unfair* and if so, in what sense? Some might concede that altering the tests in the light of boys' predilections does indeed count as a modification, and that the modified test is an improved test. Others might dispute this improvement verdict, though it is not at all clear how they could do so on the grounds that it was *unfair* to modify it. We might concede this much –that the verdict that boys really had improved their reading in comparison with girls in the year of the 'boy friendly' material is open to serious question.

Let us try again on behalf of the primary heads. If, in a particular culture, there are motivational patterns associated with boys, and others associated with girls, then is a test catering for male rather than female motivation obviously unfair? I cannot see why. It could be urged instead that differential performances linked to gender should be tackled during schooling so that the group lacking certain motivations and interests 'catch up' with the other group. (Which group should 'catch up' would depend on value judgments other than those relating to fairness).

Yet another problem with this whole approach is the arbitrariness of the categories and performance differentials picked out for concern. Gender is, of course still a high profile political issue. Nevertheless, we might find that other factors such as geographical origin, wealth or position in family also figured in how interesting children found given content.

Again, the upshot of this discussion is that conceptual problems rule out the very possibility of articulating clearly and defending convincingly some everyday verdicts about the fairness of educational assessment. This simply echoes the results of our treatment of the accommodation issues.

Unfairness in music conservatoires?

Over the last few years there have been complaints about the fact that students at Britain's music conservatoires such as the Royal Academy of Music, the Royal College of Music and others do not represent society at large. There is talk of 'under-represented groups', and pressure on these institutions to widen participation. Several conservatoires combine outreach activities, junior schools, targeted scholarships and other devices to combat the 'problem'. However, they also argue that there is really very little they can do. Aspiring musicians need the benefit of resources and cultural capital in their families from the very beginning.

The Royal Northern College is reported to offer auditions to every applicant to ensure that potential rather than achievement is assessed (Morris, 2006). This worthy aim is highly controversial. The supposed distinction between potential and achievement is particularly hard to support here, since, arguably, applicants' potential is very closely linked to their existing achievements. Their 'velocity' on application plays an important part in their subsequent musical accomplishments. (The very coherence of this point, of course depends on the possibility of making sense of 'potential', 'velocity' and so on.)

Compare this with selecting for conventional academic courses in Higher Education. Institutions such as Bristol University take account of applicants' schools' track records in addition to standard performance indicators such as 'A' level grades. They are presumably trying to identify 'potential', construed as some kind of general academic ability. Perhaps certain independent schools are too good at generating high exam scores in their students, thus successfully masking rather modest 'academic potential'.

Now surely we cannot run a parallel account for musicians. It makes little sense to suggest that some instrumental teachers have been too successful with their pupils, thus interfering with the capacity of conservatoires to identify certain applicants as having limited potential to be top-flight players. On the contrary, what actually matters in this situation is just how well conservatoire applicants play at their auditions.

We need to return to the phrase 'under-represented groups', and the implication that the conservatoires are not 'fair' to applicants in some sense. Terms such as 'elitist' and 'middle class' are often used, as if serious criticisms. If the composition of the student bodies does not reflect society at large, but is skewed towards the rich and away from the offspring of those lower in the socio-economic stakes this is held to be 'unfair'. Can such criticisms be justified?

To do so we need to find a way of understanding the justice issue here. Rightly or wrongly the conservatoires have already conceded at least two points. First, that it is appropriate to construe unfairness as failure to tap 'potential'. Second, that unfairness of this kind is at work in some fashion in the processes through which people in the United Kingdom become, or fail to become, professional musicians. However, the conservatoires need not be acknowledging a lack of fairness in their assessments of applicants. Instead they may be endorsing the criticism of the education system's role in developing musicians.

Leaving aside the practical point that there are many international students at these institutions and hence the aspiration of mirroring the social and ethnic mix of contemporary

British society makes little sense, why would it be 'fairer' if a group of putative professional violinists matched society at large? Does this apply to *all* professions and other employments? Should judges match the rest of us in terms of gender, ethnicity, socio-economic background and so on? Some have argued just this. But what about nurses, shop assistants, carers in homes for the old folk, refuse operatives, bus drivers and estate agents?

Let us turn for a moment to the long debated idea of meritocracy (Young, 1958), of 'careers open to talents' and some notion of equality of opportunity to compete for desirable employment, associated economic success and status. If we accept the 'fairness' of a meritocracy, and Higher Education's key role in channelling those with suitable merits into rewarding jobs, we might begin to generate an argument to the effect that the student mix in Higher Education should reflect society at large. A crucial assumption here is, of course that groups classified according to ethnicity or culture do not possess inherent and fixed differences in terms of 'merit' or 'potential', and that the latter two conceptions make some kind of sense.

Perhaps this was one element in the thinking of the University of California in the famous Bakke case (Oyez Project, 1978). Given the racist history of the United States, relatively few ethnic minorities found their way into prestige professions such as medicine, and hence the university felt it was 'fair' to attempt to tweak their assessment of applicants in favour of e.g. African Americans. The latter, on this interpretation of the policy, had just as much potential to become doctors as any other group, and hence whatever was needed to ensure that potential was realised should be done.

Becoming a professional musician in the UK may not be the most obvious example of economic success, yet, on some kind of meritocratic perspective, fair competition for this career might still be thought to be an important symptom of a just society. So on this view the distribution of violinists across salient categories of students in a conservatoire should, ideally, mirror that of society at large.

Many types of employment are unlikely to inspire similar arguments. Examples might include working in an old folk's home or at a sewage plant. This very point could be argued to be a profoundly depressing symptom of an *unjust* society, indeed, to constitute an *objection* to fairness construed as 'merit' winning through. Indeed, it is perfectly possible to envisage a 'fair' society embodying laws and customs in which the less attractive employment roles are distributed equitably. So a scientist might have to spend some time as a refuse operative, a company director serve as a hospital ancillary dealing with bed pans, and

so on. Le Guin (1974) explores such an idea. Such thought experiments challenge the fairness of a meritocracy.

There is another familiar problem about linking fairness to ‘appropriate’ representation of all groups in prestige professions. Critics of affirmative action policies have long pointed out the serious dangers of inappropriate essentialisms about the groups singled out for special treatment. If we are thinking of the Bakke case or others flowing from it in the USA, the families of some African Americans are comfortable middle class people with plenty in the way of worldly goods. It is ‘unfair’, or so these critics argue, to identify applicants *as members of particular groups*.

Opponents of affirmative action often say that individuals should be treated on their own merits. Moreover, individual members of favoured groups are as likely to be concerned about special treatment as anyone else. They may well wish to be admitted to Higher Education on the basis of their own talents. Note, however, that such sentiments prioritise fairness as equality of opportunity in a meritocracy. We have already recognised that other understandings of fairness are possible.

Moreover the ‘Equality of opportunity’ associated with meritocracy is often claimed to be mere ‘formal equality of opportunity’, ensuring that places in Higher Education and in high status professions are open to all applicants who satisfy clear and transparent criteria. People are supposed to succeed on ‘merit’. Such equality of opportunity differs significantly from ‘equality of fair opportunity’ (Rawls, 1971) in which people should not be disadvantaged by gender, ethnicity or socio-economic background. However, even on the Rawlsian view, ‘native talent’ is still supposed to win through, whatever the person’s background. So it is still a matter of luck – children do not deserve their ‘potential’, and hence even this might still be regarded as ‘unfair’ (see e.g. Nagel, 2003 for discussion).

Needless to say, if we allow ourselves to raise more radical doubts about the justice of a meritocratic society, then this undermines associated verdicts about fairness. Classic unease relates to concerns about the distribution of rewards and esteem among various types of employment, and indeed between those who are employed and those who are not. Some are unemployed through their own folly while others are made redundant through no fault of their own, are unable to work because of health problems, or cannot work because they have serious learning difficulties. The important point here is that equality of opportunity is not the one and only self-evident account of fairness and justice in society.

What emerges from the discussion of this section is that criticisms of conservatoires on the grounds that their recruitment is 'unfair' are not only highly contestable, but what is worse, reliant on notions of justice in society about which there is no consensus and for none of which exist definitive arguments. Moreover the standard accounts of the situation still make much use of contestable and obscure ideas of 'potential' and 'talent'. As in the earlier discussion, conceptual difficulties abound, together with unjustifiable and ambiguous notions of fairness and justice.

Unfairness in 11 plus selection

In the UK the 11 plus exam has long been argued to facilitate social mobility. Supporters point to famous achievers who rose from humble origins via their local grammar school, and who otherwise would have been 'trapped' by their social antecedents. Despite fierce criticism over many decades, some pupils still take this examination at the end of the primary stage, hoping to attend one of the remaining grammar schools. Recent research points to the examination involving 'social selection' (for instance, Atkinson, Gregg and McConnell, 2006). Middle-class children receive more coaching than their working class peers. Coaching boosts scores, and enhances chances of gaining the coveted grammar school places. Terms such as 'bias', 'elitist' and 'skewed selection' are used in connection with this phenomenon. The defence of the exam asserts or implies an association between the identification of potential and the promotion of social mobility.

First thoughts suggest that we are confronted once again with fairness as validity, an idea that we have already understood to bristle with difficulties. Test performances of heavily coached candidates do not afford valid measures of the qualities required of grammar school students. Social mobility is highlighted because it is believed that where proper account is taken of pupils' *potential* (probably, and contestably conceived of as some kind of generic intelligence in the context of the 11 plus) then pupils from poor backgrounds will sometimes join the professional classes.

However, we may wonder whether this account does justice to the concerns of the critics. Is there some sense in which they value social mobility for its own sake? Is movement from low socio-economic backgrounds into 'high status' employment thought to be a mark of a fair society? Is it felt to have an inherent value, regardless of validity questions relating to the exam? Earlier we made a passing reference to the Bakke case and noted a

meritocratic interpretation of the University of California's policy. There is a non-meritocratic understanding too. It attributes intrinsic value to a distribution of groups in higher education and prestige professions that mirrors the population at large, a value that is, on this alternative construal, held to be independent of the *potential* of the people concerned. Dworkin (2000) notes:

..our schools have traditionally aimed to improve the collective life of the community...by helping to make that collective life more just and harmonious.. p. 403

Similarly, critics of the 11-plus coaching may be concerned about the distribution of 'kinds' of people in society, at any one time and over a period of time. However, it is difficult to pin down and justify such social ideals. In the Bakke case the whole issue of race was to the fore, together with the history of the United States in that regard. Social mobility in the United Kingdom in 2008 is not about race. Why should it be considered inherently valuable for upward mobility to be promoted regardless of whether it has anything to do with 'talent winning through'? Moreover, when seeking to promote the mobility of different kinds of people, how can we justify highlighting any one way of categorising them, given that our options include geographical origin, socio-economic status, appearance, gender, sexual orientation, and religion? Perhaps different categories should have a higher profile than they do now. What about short people, those with ginger hair or those with summer birthdays? Of course, the difficulties do not end here. Evidently mobility linked to some of these categories may not be compatible with a different kind of mobility rooted in other categories.

Conclusion

Applications of 'fairness' to educational assessment reveal a rich minefield of conceptual and normative issues. This, I believe vindicates the style of philosophising from contexts and examples. Confident verdicts about what is, and what is not fair often depend on other ethical and political assumptions that are hard to justify and are in tension with each other. The value concerns that emerge from an examination of contexts exhibit features of Wittgensteinian family resemblance; resemblances between cases can be strong, but no dimension of fairness links *all* of them. Equality of Opportunity resembles, but also differs from Equality of Fair Opportunity, which in turn resonates in some ways *but not in others*, with notions of diversity and mobility in society.

The conceptual issues may be summarised as follows. I have questioned whether there is a clear division between underlying knowledge and understanding on the one hand, and its manifestation in tests on the other. Yet I also showed that this division is vital for the very coherence of fairness construed as securing the validity of assessment processes for minorities with disabilities. Moreover the notion of a construct, and whether it should be restricted to cognition rather than affect has also been shown to raise intractable difficulties. Furthermore, I was unable to defend any kind of robust distinction between accommodation and modification even though this is required if fairness is to be linked to the construct validity of tests.

Turning to the domain of the ethical and political: it emerged that judgments about the fairness of assessment hinge on verdicts about justice in society, and these, of course are deeply contested though this is rarely conceded explicitly by those bringing accusations of unfairness. Ideas of potential and talent are often assumed by proponents of fairness but they are obscure and controversial. Hence there cannot be any kind of 'self-evident' verdicts about the fairness of the assessments themselves.

Fundamental difficulties in clarifying and defending judgments about fairness rehearsed in this paper cannot be overcome by improving assessment processes. They stem from conceptual confusion and over-simplified normative judgments which lack the self-evidence their protagonists sometimes assume. None of this points to the death of educational assessment, but it strongly suggests that it should be as 'low stakes' as possible.

References

- Atkinson, A., Gregg, P. and McConnell, B. (2006) The result of 11+ Selection: An Investigation into Opportunities and Outcomes for Pupils in Selective LEAs. Centre for Market and Public Organisation, Bristol. Online at <http://www.bristol.ac.uk/cmpo/publications/papers/2006/wp150.pdf> (accessed 29th January 2009)
- Borsboom, D., Mellenbergh, G. and van Heerden, J. (2004) The Concept of Validity *Psychological Review* 111, 4, pp. 1061-1071

Bristol University (2009) 8 Assessment and Disability. Online at <http://www.bristol.ac.uk/esu/assessment/codeonline.html#disability> (accessed 29th January 2009).

Cambridge University (2008) Undergraduate Examination Allowances: Information for Candidates. Online at http://www.admin.cam.ac.uk/offices/exams/students/allowances_and_warnings_2008.pdf (accessed 29th January 2009).

Davis, A. (2009) Ian Hacking, learner categories and human taxonomies, in Ruth Cigman and Andrew Davis (eds) *New Philosophies of Learning* (Oxford, Wiley-Blackwell)

Dworkin, R. (2000) *Sovereign Virtue* (Cambridge, Mass, Harvard University Press)

Fuchs, L. S., & Fuchs, D. (1999). Fair and unfair testing accommodations *School Administrator*, 56.10 pp. 24-27

Hollenbeck, K. (1998) Determining when Test Alterations are Valid Accommodations or Modifications for Large Scale Assessment, in Gerald Tindal and Thomas Haladyna (eds) *Large Scale Assessment Programs for all Students* (Mahwah, NJ, Lawrence Erlbaum Associates Inc.)

Henry, J. (2001) Boy-friendly tests unfair, say heads *Times Educational Supplement* May 25th

Le Guin, U. (1974) *The Dispossessed* (New York, Harper and Row)

Morris, S. (2006) Conservatoires: Can they shake off their inaccessible image? *The Independent* Thursday, 23 March

Nagel, T. (2003) Rawls and Liberalism, in Samuel Freeman (Ed) *The Cambridge Companion to Rawls* (Cambridge, Cambridge University Press)

Oyez Project, (1978) Regents of the University of California v. Bakke, 438 U.S. 265 (1978). Online at http://www.oyez.org/cases/1970-1979/1977/1977_76_811/ (accessed January 29th 2009)

Pitoniak, M. and Royer, J. (2001) Testing Accommodations for Examinees with Disabilities: A Review of Psychometric, Legal, and Social Policy Issues, *Review of Educational Research*, 71. 1, pp. 53-104

Rawls, J. (1971) *A Theory of Justice* (Cambridge Mass, Harvard University Press)

Sternberg, R. (2003) *Wisdom, Intelligence, and Creativity Synthesized* (Cambridge, Cambridge University Press)

University of Brighton (2009) Disability and dyslexia Examinations and assessments. Online at <http://www.brighton.ac.uk/disability/assessments.php?PageId=352> (accessed January 29th 2009).

University of Greenwich (2009) Guidance for Lecturers Marking the work of Students with Dyslexia. Online at <http://w3.gre.ac.uk/students/affairs/dd/staff/marking.htm> (accessed January 29th 2009).

University of Wales (2006) EXAM PROTOCOL: Additional provision and support for Disabled Students and Students with Dyslexia. Online at http://www.uwic.ac.uk/disability/students/Exam_protocol_students.htm (accessed January 29th 2009)

Wittgenstein, L. (1953) *Philosophical Investigations* (Oxford, Blackwell)

Young, M. (1958) *The Rise of the Meritocracy* (London, Thames and Hudson)