



This is a repository copy of *Cultivating the myopic learner: the shared project of high and low-stakes assessment*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/43705/>

Article:

Allen, A. (2012) *Cultivating the myopic learner: the shared project of high and low-stakes assessment*. *British Journal of Sociology of Education*, 33 (5). pp. 641-659. ISSN 0142-5692

<https://doi.org/10.1080/01425692.2012.668832>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Cultivating the myopic learner:

The shared project of high and low-stakes assessment

This paper argues that despite the obvious and important differences between high and low-stakes assessment, there remain important points in common. These manifest themselves at a sociological level, where each tradition of assessment shares a similar disposition towards power. It is argued that both high and low-stakes assessment as they are practised in England today, act together in support of a wider regime of power. This regime relies upon the construction of specialised subjectivities, defined by a myopic and self-perpetuating concern with individual progression.

Keywords: Assessment, biopower, disciplinary power, process.

Ansgar Allen

The School of Education

The University of Sheffield

388 Glossop Road

Sheffield

S10 2JA

a.allen@sheffield.ac.uk

Word Count: 7962

Biography: Ansgar Allen's research is in the areas of educational philosophy, history and theory. It traces the interrelated histories of schooling, meritocracy, and assessment, and relates these histories to trends in biological, statistical and political discourse along with broader shifts in regimes of power.

Cultivating the myopic learner:

The shared project of high and low-stakes assessment

This paper argues that despite the obvious and important differences between high and low-stakes assessment, there remain important points in common. These manifest themselves at a sociological level, where each tradition of assessment shares a similar disposition towards power. It is argued that both high and low-stakes assessment as they are practised in England today, act together in support of a wider regime of power. This regime relies upon the construction of specialised subjectivities, defined by a myopic and self-perpetuating concern with individual progression.

Keywords: Assessment, biopower, disciplinary power, process.

Introduction

It is common to oppose 'high-stakes' testing in schools to a 'low-stakes' alternative as if these two forms of assessment represented two distinct logics. High-stakes testing is dominated by demands for accountability, accreditation and scientific rigor. Low-stakes testing, by contrast, has a reputation for being softer and less judgemental. It is concerned with the immediate learning needs of the child and often makes use of interpersonal techniques. There is a convention in educational circles which is to maintain that there is too much quantitative testing in schools, that the examinations system is overbearing, that it warps teaching practice and invades pupil subjectivities, that it even defeats the very point of education (Thomas 2005; Siegel 2004; Hall et al. 2004; Reay and Wiliam 1999). In this climate, any move towards more pedagogic, qualitative and seemingly humane forms of assessment, appears to be fighting against this trend.

This paper argues that despite the obvious and important differences between high and low-stakes assessment, there remain important points in common. These manifest themselves at a sociological level, where both traditions share a similar disposition towards power. This is important because low-stakes assessment is often offered as the solution to high-stakes testing. Some commentators believe that high-stakes testing in English primary schools and in the first years of secondary schooling may be on the brink of substantial reform (Whetton 2009). In this context, Wyse and Torrance (2009) argue for a replacement system that combines light sampling of schools, for accountability purposes, with an increased emphasis on formative assessment,

geared towards learning. The hope amongst researchers such as these is that the burden they associate with high-stakes testing will be removed. This paper questions the foundational assumptions upon which this debate rests and asks: to what extent from the perspective of power, would the reduction or removal of high-stakes testing have an impact? To answer this question, I consider in turn the recent histories of high and low-stakes testing as they relate to the question of power.

High-stakes assessment

High-stakes testing in schools has a long history that for reasons of space cannot be considered here. It is sufficient to note that towards the end of the twentieth century, testing in English education reached unprecedented levels in terms of frequency, scope and uniformity. A major landmark was the 1988 Education Act which established a national curriculum. This was accompanied by periodic national tests (widely known as SATs) administered at 7, 11 and 14 years. High-stakes testing was embedded throughout schooling in a coordinated scheme of examinations and associated pupil-tracking technologies.

The development and implementation of SATs deviated substantially from the original blueprint provided by the Task Group on Assessment and Testing (TGAT). TGAT attempted to combine teacher-led formative assessment with external demands for accountability (DES 1988). In the early 1990s, teacher-led assessment was quickly abandoned in favour of external, nationally comparable tests. Some commentators felt that this was inevitable (Kimberley et al. 1989; Troman 1989; Gipps 1992; and Torrance 1992). According to its critics, the original blueprint provided by TGAT contained basic flaws and contradictions arising from its attempt to reconcile divergent aims. The blueprint's dual purpose, provide accountability/assist learning, could only result in irreconcilable tensions; a classic case of high versus low-stakes testing.

The main architect of this system, Paul Black, offered a different analysis, claiming that the original TGAT blueprint was distorted (Black 1994, 1997, 1998). A series of changes were made to the system of testing which were not 'informed by any overall strategy' (Black 1997, 45). The two decade history of SATs is a tale of multiple, conflicting interests (Whetton 2009). Without wishing to contradict these accounts, I nevertheless believe that it is possible to subject the events they recount to a different level of analysis. I argue that alongside the multiple

interventions and adjustments that make up the recent history of SATs, can be found a guiding framework of power which has been far more coherent and persistent than is suggested by these histories of distortion and competing interests.

Disciplinary Power

At first sight, the proposals outlined by TGAT seem to resemble a regime of the sort once described by the philosopher Michel Foucault. In his study of prisons, schools, barracks and hospitals at the turn of the nineteenth century, Foucault observed a set of techniques that he identified with an emerging form of ‘disciplinary power’ (Foucault 1975). The appearance of modern examination can be located at the core of these developments (Hoskin 1979).

Towards the end of the twentieth century, TGAT did not, however, recreate some sort of early nineteenth century disciplinary regime. The system of high-stakes testing that arrived through subsequent distortion and adjustment had its own independent coherence, which owes more to the systems of power in operation at the turn of the twenty-first century, than it does to systems of disciplinary power found over two centuries ago.

The original intention of TGAT was to embed a detailed machinery of ‘criterion referencing’ throughout the interim years of schooling. The ‘absolute quality of the performance’ of an individual child was to be assessed periodically alongside an explicit set of criteria (DES 1988, 4). According to this scheme, the developmental process of a “normal” child would be divided into parallel components or *elements* (see figure 1) that were themselves to be subdivided into *levels*. Every level was finally to be divided into a set of specific *targets*, which would match the prescriptions of the national curriculum. Assignment of targets was to take place in context, in the classroom, and was to be carried out by the teacher (DES 1988). Testing would allow for a judgement to be made about a child’s position in a pre-defined sequence of progression enabling the teacher to provide developmentally specific feedback to the child.

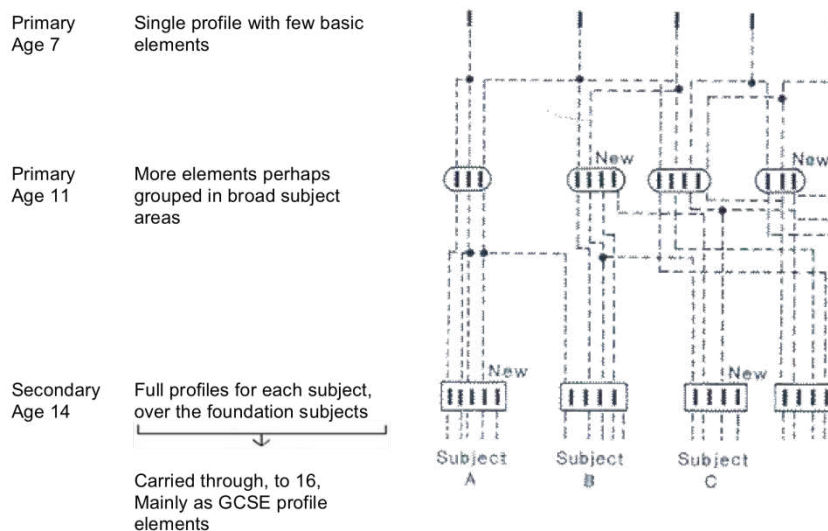


Figure 1. Location of child according to the development of 'elements'.

Each element is divided into 10 levels. Every level is divided into targets.

(DES 1988, 38)

This sort of process, where a child is precisely defined according to the place he or she has reached in a prescriptive series, is recognisably disciplinary. The child is located in 'cellular' space. There is a division of schooling into 'successive or parallel segments' (Foucault 1975, 157). Progression is charted through periodic examinations, ensuring that all follow the same course. As a result of these measures, power invests a 'duration that would otherwise elude one's grasp' (Foucault 1975, 160).

Nevertheless, and despite these superficial affinities, this system is not disciplinary in the original sense. Crucially, the developmental scale employed by TGAT contains a set of assumptions concerning what the *average* child is expected to achieve at each respective level. As critics soon observed, the proposal did not only make use of criterion referencing, it also involved *norm referencing* (Croll 1990; Wiliam 1992). This is clearly evident in figure 2 where the linear progression of a normal child is depicted. The bold line gives the expected results for an average child at the ages specified. The dotted lines represent the limits within which about 80 per cent of non-average children may be found. This calculation presumes a normal distribution of ability in the population at large, which means the 'disciplinary' mechanisms of power outlined above are at the very least framed by a population-level awareness. It should be noted that the normal distribution was a mid-to-late nineteenth century invention (Porter 1986). It came after the period considered by Foucault (1975) in *Discipline and Punish*, and is associated with a very different form of power, a type of power Foucault later called biopower.

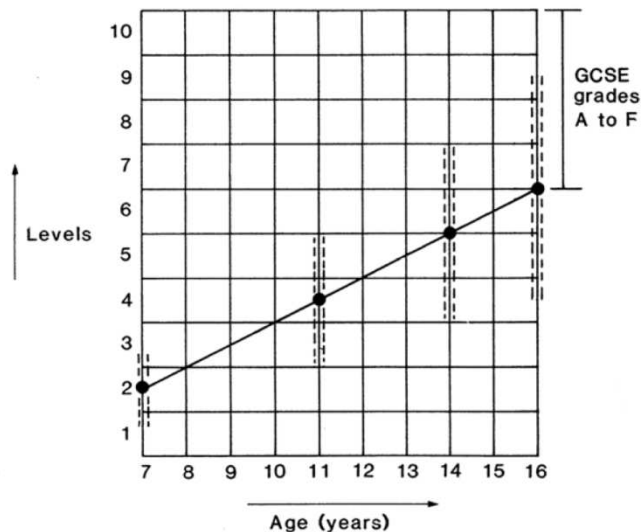


Figure 2.
Normalized sequence of progression
 (DES 1988, 33)

Biopower

The concept of biopower was developed by Foucault (1976a, 1976b, 1978) shortly after his work on discipline. Biopower and disciplinary power represent two very different approaches to the problem of social governance. From the perspective of biopower, society is viewed as an organism with natural cycles and processes that must be discovered and then nurtured. The point is to act *with* existing relations in mind. Disciplinary power, by contrast, works *on* the population it subjects without due regard for its pre-existing forces and tendencies.

When outlining his concept of biopower, Foucault was careful to explain how it was established *after* the development of disciplinary technique. Recapping his earlier work on discipline, Foucault explains how these earlier techniques were ‘essentially centered on the body’ (Foucault 1976a, 242), which they sought to separate out, align, serialise and make visible. Discipline manipulated the body ‘as a source of forces that have to be rendered both useful and docile’ (Foucault 1976a, 249), whilst employing ‘techniques for rationalizing and strictly economizing on a power that had to be used in the least costly way possible’ (Foucault 1976a, 242). This relied upon a careful distribution of disciplinary power throughout an extended network of relations that helped to spread and reduce its weight.

Biopower was more complicated to implement and this is why it took hold somewhat later. Whilst discipline was developed within an institutional setting, biopower required ‘complex systems of coordination and centralization’ (Foucault 1976a, 250). It operated at the level of populations, and this depended upon systems of knowing and intervention that transcended the

institutional boundary. Unlike discipline, which works by manipulating ‘man-as-body’, biopower applies itself to ‘man-as-species’ (Foucault 1976a, 242). Biopower is addressed to ‘a global mass that is affected by overall processes characteristic of birth, death, production, illness, and so on (Foucault 1976a, 242-3).

Looking back over his earlier treatment of discipline, Foucault decides that a switch in terminology is required. What is involved in disciplinary technique is actually a process of ‘normation’ rather than ‘normalization’ as it was earlier labelled in *Discipline and Punish* (Foucault 1978, 57). Discipline is distinct from biopower because it begins with a norm. It ‘breaks down individuals, places, time, movements, actions, and operations’ and attempts to recombine them according to its prior model (Foucault 1978, 56).

Biopower adopts a different approach, beginning with the measurement of existing phenomena and deriving norms from the results obtained. With biopower, the ‘normal comes first and the norm is deduced from it’ thus ‘what is involved here is no longer normation, but normalization in the strict sense’ (Foucault 1978, 63). Biopower works backwards from the discovered traits of a population and the normal distribution they reveal. This process can be observed in the TGAT blueprint, to the extent that it makes implicit use of norm referencing and thereby appeals to the concept of a normal distribution of abilities. Therefore, the TGAT blueprint was not purely disciplinary as it demonstrates an awareness of population level traits. This is a theoretical point. However, it alerts us to a new set of possibilities: SATs had the potential to develop in a biopolitical direction.

Process rather than position

Disciplinary power treats any large population as if it were a mysterious entity. The population is a formless mass of bodies from which individuals are extracted and then put to use. For biopower, by contrast, the population is a cognisable social organism; it is an independent entity that contains ‘a set of processes to be managed at the level and on the basis of what is natural in these processes’ (Foucault 1978, 70). The population is no longer an aggregate of individual bodies that are to be controlled on an individual basis. There are certain collective phenomena that ‘become pertinent only at the mass level’. These phenomena are ‘aleatory and unpredictable when taken in themselves or individually’ but ‘at the collective level’ they ‘display constants that

are easy, or at least possible to establish' (Foucault 1976a, 246). The desire to pull the causal strings of a deterministic world is, therefore, replaced by a very different view of government, one that concerns itself with the 'taming of chance' (Hacking 1990). 'The population is a datum that depends upon a [more general] series of variables' that are not immediately transparent. This means that 'the relation between population and sovereign cannot simply be one of obedience or the refusal of obedience, of obedience or revolt' (Foucault 1978, 71). Instead, there is a relation between population and government that is one of *management*. The point is to define 'a bandwidth of the acceptable that must not be exceeded' and then allow natural processes to unfold within this boundary (Foucault 1978, 6). Techniques designed to cope with these natural processes now come to define the basic repertoires of government.

The subsequent development of SATs can be seen within this framework as moving from the neo-disciplinary, and as we have seen, partly anachronistic TGAT blueprint, towards a well-established biopolitical regime of testing. This transformation has, nevertheless, been hard to perceive as biopolitical assumptions are well established in educational discourse, restricting the perceptions of critique. The early critics of TGAT were consequently unable to perceive the surrounding regime of biopower that would soon modify the TGAT proposals (their mistaken target was disciplinary power). As Torrance argued in 1989, the initial TGAT scheme 'betrays a highly instrumental and mechanistic view of learning' for which 'teachers are simply taken to be operationally responsible' (Kimberley et al. 1989, 245-6). Such approaches to learning may have worked during the early nineteenth century when disciplinary power was first developed, but according to Torrance, writing in the late twentieth century, it appears 'too mechanistic'. Education is now seen as 'unpredictable' and not amenable to such approaches (Kimberley et al. 1989, 247). In the late twentieth century, it is still possible to design highly mechanistic technologies as TGAT has shown (DES 1988), but critics are quick to mention that such technologies are unlikely to function as planned. This is because the "nature" of the educational task is now seen as far too complex. And yet, according to my argument, biopower took hold in late twentieth century schooling, *precisely because education is now seen in this way*.

From this commonly adopted perspective, one which affirms the irreducible complexity of human life, it is assumed that a single route of progression cannot adequately represent the

learning process, however intricate and well-branched its architecture may be. Assessment must now embrace the fullness of human life and allow bodies their own space to move. Therefore, the prescriptive route of progression found in the 'TGAT' blueprint led, through various distortions, to a technique that was profoundly non-disciplinary in its form. The result was a mode of assessment that operated according to a loosely normalized framework of progression.

Within this biopolitical framework, the idea of *process* is now more important than *position* (where establishing position was the first and fundamental requirement of disciplinary power). We thus move from the accurate recording of ability and careful positioning of individual learners found in the 'TGAT' proposal, to the more flexible supervision of naturally occurring movements. This takes us from a type of disciplinary machinery where docile bodies are efficiently positioned in space according to their respective merit and utility, to a system where bodies must be encouraged to position themselves heedless of the overall distribution. Whether they are positioned perfectly or not, is no longer significant. When the complexity of nature is embraced, the *act* of positioning is more important than the *accuracy* of the position attained. Where disciplinary power seeks to define exact location and then the precise movement of such located bodies, biopower is concerned with the support and management of the natural processes it perceives.

Black (1998, 57) concluded, a decade after 'TGAT' convened, that 'current national testing in England and Wales falls far short of acceptable requirements of reliability and validity'. From a technical perspective, this statement is straightforward enough, it merely reports on the "malfunctioning" of testing machinery. However, from the biopolitical perspective outlined above, this is symptomatic of a shift in strategies of power and could even be judged a success. Assessment is no longer a disciplinary technology that helps to micromanage learning, rather, it is a technique that provides an overall framework for interaction. Achieving high levels of reliability and validity is no longer important because, from the perspective of power, there is little need to establish the precise location of the learner with a pre-set scale of progression.

As the original 'TGAT' blueprint broke down, Standard Assessment Tasks became summative written tests, teacher assessments were downgraded in favour of SATs results, and marking became external. 'Targets' came to serve the purpose 'TGAT' planned for 'elements',

and the careful subdivision of elements into levels became confused. Overall, the system was undergoing radical simplification (Black 1997). According to Black and Wiliam (2003, 625), all that remained was the ‘idea of age-independent levels of achievement’ and regular examination at the end of each key stage. It was impossible for the learner to know his or her precise location within an explicit route of progression

After these changes, distortions and simplifications, all we are left with is the disembodied idea of process. That is, we are left with an idea of process that has been stripped of any precise notion of where one is within a sequence of progression. What remains is a near-sighted or myopic view of individual enhancement, of improvement abstracted from any detailed framework of progression or overall plan. Here, we have process without precise location, without informative outcome: process existing for its own sake, for the sake of proceeding. Process is prioritised as a natural sequence of changes without independent purpose or end.

Value-added and process monitoring

The biopolitical preoccupation with process can be found within the original TGAT recommendations, despite the more prominent though soon-to-be neglected emphasis on position. This can be observed in proposals to embed ‘feedback’ and ‘feedforward’ techniques (DES 1988). It should be noted that *feed* is a process term, meaning to supply material progressively for an operation, *but also including the means, channel, motion or rate of such supply*. My argument is that, within the domain of biopower, the act of movement itself (its motion, means and rate) has had a tendency to become the point of interest rather than the material being dealt with or the eventual output or endpoint.

The implementation of *feedforward* techniques was initially more successful. These techniques seek to bring about the modification or control of a process using its anticipated results or effects. According to TGAT, feedforward would occur when data is passed on during class-to-class transfer, and alongside transfer at eleven from ‘feeder schools’ to secondary education (DES 1988). The rise in feedforward techniques is part of a much wider expansion over the past two decades in the availability of educational data (Schagen and Hutchison 2003). More widely available, nationally comparable datasets allow for deeper statistical analyses. This would enable the greater pursuit of phenomena that ‘become pertinent only at the mass level’

(Foucault 1976a, 246), and would encourage the development of associated pedagogic techniques designed to deal with these aggregate effects.

One product of such data production has been the calculation of ‘value-added’ between periodic tests. Value-added is defined as the difference between a statistically predicted performance and the actual performance. A score of zero denotes average progress, while scores above zero indicate above-average progress[1]. The calculation of value-added builds upon the assumed sequence of progression first developed by TGAT. According to its 10 level scheme, an average pupil aged 8 would be performing at the lower end of level 3; an average 10 year old should have achieved level 3; a 12 year old level 4; and a 14 year old level 5 (see figure 2). Following statistical analyses of the progressions of actual children, a *likely-future* is predicted for each child type according to his or her starting score. This allows schools to be judged by parents and government, and teachers to be judged by school administrators, according to their ability to achieve (or exceed) the likely-future expected for each of their pupils. Those that enhance individual development (beyond what may be predicted by extrapolation from normalized progression routes) would have generated surplus value.

In a system of value-added, when process itself becomes the predominant point of interest, the *rate* of progression is the focus for improvement. Individual or institutional performance at any one time is still relevant but what exactly that performance represents in terms of what a pupil is able to do when tested is becoming less significant. Whether a level 5 in Mathematics at Key Stage 2 in 2002 is equivalent to a level 5 in Science at Key Stage 3 in 2005, no longer matters in a value-added system so long as ‘all schools have taken the same tests’ (Ray 2006, 18). The percentage of pupils achieving five A* to C grades at GCSE today serves as a supreme benchmark of institutional performance. Yet, the aim, simply, is to increase this statistic.

In contrast to the disciplinary goal, which is to establish the position of individual bodies, define their functional capacities, and then put them to work within a larger machinery, the focus here is less specific. The aim is to enhance individual improvement, *as a life-force to be encouraged for its own sake*. Biopower, it should be recalled, aims to support what are perceived to be natural processes, simply because they are said to exist. The exact meaning of a specific outcome is no longer significant once measures of temporal difference take over. In other words, when the

distance between two points is the central concern, what these points represent in themselves matters less.

The publication of value-added statistics took time to implement. Cohorts of students would first have to pass through the national curriculum framework and its associated tests. Only then could sufficient longitudinal data become available to calculate student trajectories and generate progression norms. However, the switch to value-added technologies has not been without its critics. Some researchers have attempted to ‘counter the recent overwhelming focus on ‘process’ which, they believe, masks deep structural failings of the education system (Gibson and Asthana 1998, 271). It is felt that when research becomes exclusively concerned with intra-school processes it tends to neglect the effects of broader structural factors. Ironically, a process-based solution to such critique later appeared. Subsequent researchers found that social disadvantage has a statistically negative impact not only on outcomes, *but also* on the potential for both school improvement (Levacic and Woods 2002) and individual progress (Feinstein 2003). This demonstrated that educational process as well as educational outcome is affected by social factors.

As a measure of educational process, value-added systems had at first ignored the effects of social inequalities. They only took account of prior attainment, ignoring other external influences on performance. In response to criticism, value-added systems were later contextualised to incorporate structural factors, such as gender, ethnicity and social disadvantage. Gross population norms were now divided into specific norms that were appropriate for sub-population groups. Likely-future predictions were now served by contextualised norms.

Contextual value-added (CVA), as it came to be known, required more sophisticated analysis, finer discrimination of the population and more data. In 1999, ‘unique pupil identifiers’ were introduced allowing successive key stage scores to be matched to individual pupils, all of which could be tracked through a central database. In 2002, the Pupil Level Annual School Census (PLASC) commenced. This required schools to record and submit background characteristics for each pupil. With these enhancements to the data stream, the likely-futures for contextualised children could now be calculated.

The calculation of value-added is a statistical operation, designed mainly to inform, allowing individuals and institutions to subsequently govern themselves. The expectation is that school managers, teachers and pupils will adopt the principles of process enhancement and translate these principles into contextually specific efforts to achieve individual gains. This approach is definitive of biopower which functions alongside existing tendencies and situations, seeking to intervene through remote influence at the level of lived experience. Nevertheless, some schools are thought to require additional motivation: In 2008, a government initiative called *Gaining Ground* was launched to target so-called 'coasting schools' where the flow of pupil progress since leaving primary schools is demonstrably reduced. Their pupils suffer from a loss of 'momentum', which is a process-based malaise (DCSF 2008, 1-3). Local authorities must identify coasting schools and provide help and resources, including recruiting advisors from more process-oriented schools in the area.

The *Gaining Ground* initiative provides examples of prototypical process-enhancing schools. School strategies include so-called traffic light warning systems, monitoring cycles, regular progress reports to parents, scrutiny of teaching practice, mentoring and targeted interventions, all geared towards the demand for improvement as measured by a 'rigorous target-setting and data tracking system' (DCSF 2008, 11). In a separate government report, one school is described in which pupils carry 'a regularly updated grid in their planners, downloaded from the system, indicating their current attainment and personal targets'. This is combined with regularly updated cohort lists 'posted near the school canteen, showing students' current levels' (Ofsted 2009, 25).

Whilst these trends may appear to be overbearing and intrusive it would be a mistake to describe them as disciplinary. It is only the momentum of human life and learning that is to be targeted here. How greater momentum is to be achieved is left open to localised techniques. The overall strategy is, nevertheless, concerning as it acts to sustain and intensify a myopic preoccupation with the processes of individual enhancement.

Pupil tracking and the prediction of likely-futures

The above school-based strategies for process enhancement depend upon well developed pupil-tracking technologies. Long before value-added statistics were made available for general consumption, systems were introduced to schools for purposes of internal management.

Initially, this depended on university-based research groups offering value-added analysis as a service to schools (Schagen and Hutchison, 2003), though outsourcing has since been superseded by freely available interactive software. *RAISEonline* is one such facility enabling schools to access their results and interrogate performance data in greater depth. School, group, pupil and question level data can be analysed; comparisons can be made to national patterns; performance in specific curriculum areas can be monitored; and individual pupil targets can be set, informed by progress made by similar pupils nationally. The software is interactive, allowing users to import extra data from optional tests, compile customised reports and even experiment with hypothetical futures. Thus, 'school administrators are able to amend and change data, such as adding and removing pupils or changing grades' allowing them to carry out 'what if scenarios' (RM Data Solutions 2011). Possible interventions can be weighed against their likely-future effects.

Process calculations based on national tests have been supplemented by additional tests that compete with SATs in terms of their predictive power (Strand 2006). For example, a large proportion of the school population completes the Cognitive Abilities Test (CAT) shortly after arriving at secondary school. This testing package comes with the option to make use of its detailed predictive reports, which can be used for purposes of internal management or at a year 7 parents evening where a child is given his or her predicted grade in, say, GCSE History. Predictions such as these relativise the child within his or her own likely-future, a future that is the product of statistical manipulation and population norms. The child is encouraged to operate within this space of zeroed progress, aspiring to above average development, to the hope of positive-value added, towards a surplus value that becomes his or her maximum hope.

The effects of monitoring upon pupil subjectivities have not escaped research interest. In a study of pupil self-perception, Reay and Wiliam (1999) explore the effects of high-stakes testing. They argue that notions of self-worth for year 6 children are linked to the level they hope to achieve in upcoming SATs. The problem with these tests, we can infer from this study, is that they reduce complex selves to single scores. Take Hannah, for example, who, fearing poor performance, thinks she will "be a nothing". However, as the authors note, she 'is an accomplished writer, a gifted dancer', artist and so on, 'yet none of those skills make her

somebody in her own eyes' (Reay and Wiliam 1999, 345-6). These qualities, they argue, are missed by SATs, which produce 'very simplistic judgements purged of any subtlety and complexity' (Reay and Wiliam 1999, 349). Several years later Hall et al. (2004) reached similar conclusions. They refer the production of 'amenable pupil identities' in an environment where organisational instructions 'are unspoken', pupils are 'well-drilled' and 'police themselves' in lessons dominated by preparation for upcoming tests (Hall et al. 2004, 806). In both of these reports, we can see how critics rely upon the construction of a binary between reductive or simplistic SATs and 'learners who have complex, multiple identities' (Hall et al. 2004, 802). Monitoring systems are portrayed as overly mechanistic and disciplinary.

These findings are important, but subsequent developments in value-added technologies have led to more subtle adjustments designed to build upon (rather than deny) the complexities of lived experience. This is why defending education against mechanical reduction is becoming increasingly ineffective as a challenge to current strategies of power. Pupils are now located along their own normalised scheme of progression reflecting the average traits of their sub-population group. These personalised lines of advance have been statistically adjusted to the complexities of real lives. This 'developmental' scheme is not organised according to the disciplinary management of time where rank and position are the most important factors. Equally, there is no attempt to place tested and docile bodies within a larger functional architecture. The strategy of subjection is far more open and individualistic, *embracing* complexity in a way that is missed by the research studies summarised above. *Progress itself* is sought, a form of progress that is narrowly defined of course, but which has been abstracted from the detailed, subdivided scheme of advancement towards which disciplinary power aspires. This is not to say that single test outcomes are no longer of great personal import or that ranking ceases to exist, but it does suggest a supplementary logic. Pupils are situated so that their complex selves are able to unfurl in a personalised framework. Assessment technologies have opened up to complexity, and in that sense they are developing in a profoundly *non-disciplinary* direction.

Pupil tracking now begins at a very early age. Since 1998, when the systematic assessment of all children entering English primary schools became a statutory requirement, various techniques have been developed. These include 'objective' baseline tests that make it possible for

schools to predict likely-futures from 5 years of age, monitor individual trajectories, and devise strategies that maximise progress throughout the primary phase. In its mission statement, one of the major providers of monitoring systems offers a good account of its work. They apply, we are told, ‘the best standards of science’ to ‘keep things ‘in perspective’ (Centre for Evaluation and Monitoring 2011). This last point is essential. The whole purpose of educational process monitoring is to generate perspective. To get things in perspective means to reduce things in magnitude so that they appear realistic and manageable, in this case, to generate pupil specific expectations of future performance.

Process-supervision and perspective-adjustment are features of contemporary educational practice that are not exclusive to high-stakes testing and its associated statistical techniques. In the domain of low-stakes testing, as we will see, very different approaches are used in pursuit of similar aims.

Low-stakes formative assessment

According to the 1988 TGAT blueprint, national assessment would incorporate a process of *feedback*. Advice would be provided in response to the output of a child’s work, returning the output of a system as an input, as a means for improvement and self-correction. TGAT made the mistake, according to its early critics, of portraying this process in an overly mechanical way. In the scheme they devised, feedback was designed to refer to an externally set sequence of progression that had been divided into elements, levels and targets. As we have seen, this highly mechanised approach was largely abandoned from the developing system.

Years later, assessment-led-learning was reintroduced as an exclusively low-stakes endeavour, called ‘formative assessment’ where the external validity of a teacher’s feedback was no longer important (Black and Wiliam 1998a, 1998b). Rather, the response offered by the pedagogue (or peer) was to be impressionistic, only required to make sense in the context within which it is made.

Formative assessment is characterised as a cycle in which evidence of an incidental, ephemeral and continuous sort is used. The near-sighted aim is to close ‘the gap between actual and desired levels of performance’ in a self-referential loop (Wiliam and Black 1996, 543). The technique is interpersonal and relies upon the construction of relationships rather than statistical

constructs as above. Nevertheless, formative assessment can be thought of as *building on* the techniques already outlined, extending process-based strategies further into pupil subjectivity and to a finer level of detail. The specific context of each learner is to be prioritised and penetrated by pedagogic technique, where an individual's unique developmental path becomes the site of manipulation. As with the techniques of high-stakes testing, formative assessment relies upon a comparison between one's position today and what one might be tomorrow. This, too, results in a shortened focus, one that has become absorbed in the process of learning itself. It is a game of ceaseless self-monitoring and self-enhancement under the sign of 'growth', one that maintains subjects in a state of constant activity through its mechanisms of myopia.

It is interesting to explore the narrative within which this work is situated. We are told that, historically, the 'day-to-day activities of teachers' have been 'of little interest to academic researchers' in assessment (Wiliam and Black 1996, 537). In the 1990s this changed, when Wiliam and Black journeyed into the "black box" of the classroom, into a void of unknown relationships neglected until then by a type of assessment research that had been preoccupied with simple input/output variables. Though the discovery of formative assessment is recent, its existence as an undiscovered activity is presented as timeless, presumably to be found in all good teaching throughout the centuries. It is 'nothing new', we are told. Almost 'all successful teaching' relies on adapting approaches 'in the light of evidence about the success of previous episodes' (Wiliam and Black 1996, 538). In order to be enhanced, this form of assessment must be conducted in an environment that is carefully managed. The surroundings must be safe and comfortable for the student, allowing learning to flourish. 'Good teachers know that this is so' (Weeden et al. 2002, 16). The caring and carefully responsive type of teacher is the sort that must be modelled and enhanced. As a result, research aiming to improve formative teaching practice, must not be prescriptive, it should work with what teachers already do and 'build on existing good practice' (Black and Wiliam 1998b). Formative assessment appears as an already existing *natural* pedagogic process that can be enhanced or indeed hindered by educational policy.

It was argued above that process is now more important than position. However, in formative assessment the supervision of process becomes an ideal in itself. Indeed, it is said to enhance self-esteem and remove the damaging effects of ranking and competition amongst

peers. 'Where the classroom culture focuses on rewards, 'gold-stars', grades or place-in-the-class ranking, then pupils look for the ways to obtain the best marks rather than at the needs of their learning'. This also generates a 'fear of failure' and leads to efforts by pupils to 'try to build up their self-esteem in other ways'. What 'is needed is a culture of success, backed by the belief that all can achieve' (Black and Wiliam 1998b, 8-9). Confrontations with failure should be avoided, and the idea of failure itself must be suppressed beneath a process of constant optimistic striving.

"Helpless children" who 'seem to accept that they will fail because they are not clever enough; believe that if something seems to be too hard there is nothing they can do about it'. They 'tend therefore to avoid challenge' (Weeden et al. 2002, 53). These children must be transformed into "mastery children" who 'are motivated by the desire to learn; will tackle difficult tasks in flexible and reflective ways; are confident of success, believing that they can do it if they try'. Mastery children 'believe that you can improve your intelligence' (Weeden et al. 2002, 53). The transformation is not easy, and such children may have to pass through an intermediate phase on the way from learned helplessness to mastery. In this transition phase 'they find strategies that place the blame on others or factors beyond their control' (Weeden et al. 2002, 54). Transition will occur by breaking down learning into the minute and achievable steps of formative improvement and by encouraging pupils to take a greater responsibility for their learning instead of extending a critical gaze elsewhere.

The deliberate strategies adopted by the formative teacher to avoid scoring individuals, along with efforts to create a 'culture of success' in the classroom, are also part of this governmental shift. But formative assessment involves far more than teacher feedback. It requires that teachers instruct pupils in the techniques of formative assessment itself. In peer-assessment, pupils engage in mutually beneficial feedback conversations that allow them to assist each other in achieving individual advancement. 'Individual pupils learn how to communicate with their peers in non-judgemental ways' soon discovering that 'if they want constructive feedback they have to be sensitive about the kind of feedback they give others' (Weeden et al. 2002, 89). The assessment process is socialised and rendered humane as pupils learn to work in an environment of individual striving, without the deleterious symptoms of rivalry. Pupils are to be trained in the techniques of cooperative yet atomised activity.

Self-assessment is considered to be at the core of formative assessment allowing pupils to develop a constructive and active approach to their own development. Along with formative teaching, we are told that self-assessment is nothing new, in fact, it is an entirely natural process particularly prevalent in the better learner: ‘many high attainers already intuitively self-assess’ (Weeden et al. 2002, 76). However, these processes of self-development are not to be left free from constraint. Crucially, for self-assessment to be effective, pupils must be told what they are expected to achieve, they must be made aware of their learning goals. This is a display of professional openness on behalf of the teacher that appears intrinsically democratic.

There can be a certain degree of negotiation and reformulation of goals, which helps to ease their internalisation, but ultimately these learning goals are translated from external sources. One pioneer teacher refers to ‘pupils taking *ownership* of their own learning’, how she lets them ‘know what the intention of the lesson is and what they need to do to *achieve it*’ so that they can take ‘*responsibility* for their own learning’ (Black et al. 2003, 49). These three terms emphasised in the original text - ownership/achievement/responsibility - sound empowering. However, they disguise a process by which goals are transmitted to the student for subsequent internalisation. Though such an approach may well be pedagogically effective and may be empowering to a degree, there is a subtext here. The deeper agenda of self-assessment is to develop the ability in subjects to enhance personal performance *within a logic that is externally set*. Pupils learn to perform work on themselves according to an external design. This is typical of biopolitical supervision, the point of which is to define ‘a bandwidth of the acceptable that must not be exceeded’ and then allow so-called natural processes to unfold within this boundary (Foucault 1978, 6).

In the formative classroom, the pernicious presence of competition is to be removed. However, the removal of competition is more apparent than real. The point is to cultivate an environment where competition, a supposedly natural human tendency, is allowed to unfold without negativity. Competition must be positive and productive. In the formative classroom pupils are taught to compete through processes of self-analysis and self-enhancement. They learn to assess their current position and make adjustments towards a goal that is considered realistic and within reach. Despite claims to the contrary, competition does not disappear within the explicitly constructive and encouraging ethic of formative assessment. What retreats is the

goad of ranking. In its place pupils learn how to enhance process and develop themselves in apparent harmony with one other, each of them involved in personal formative cycles, occupied in unison within individual feedback-action loops. They learn to become industrious self-enhancers, accepting and implementing external goals. Competition is humanised and disguised and perhaps thereby intensified by this formative technology.

Conclusion

Both high and low-stakes testing as they have been developed in recent years demonstrate a similar disposition towards power. High-stakes testing is commonly associated by its critics with mechanistic, disciplinary and reductive techniques. However, this ignores the extent to which it has become connected to process-oriented pedagogies. These seek to embrace the complexity of individual development by referring to each learner in probabilistic terms, in terms of his or her own likely-future. In a similar way, low-stakes formative assessment concerns itself with the immediate environment of the child. Learners are to work together in cooperative rivalry, using comparison with others as a vehicle for reflecting on their own development. Overall, these techniques have been formed out of a common biopolitical assumption that learning itself is a natural process, requiring instruments and techniques of assessment that will support it and magnify its potential. As its own justification, biopower leads to the intensification of such practices that are concerned with self-fabrication and self-enhancement. Learning in a biopolitical regime of high and low-stakes assessment is both naturalised and contrived in a way that is starkly individualistic.

How one might resist such a system remains an open question. From the above argument, I can only suggest what resistive activity should avoid or at least be wary of: (1) It should not be assumed that low-stakes testing will help counteract or reduce the systems of power that are tied up with high-stakes testing. (2) One cannot rely on appeals to the irreducible complexity of the individual learner, as a means of defence. Indeed, such appeals have the potential to contribute to the systems of subjection that they attempt to subvert.

Notes

1. The value-added system was later zeroed arbitrarily at a score of either 100 or 1000 to avoid negative scores.
2. The future of the CVA system is now in doubt, following a change of government and a promise to remove CVA as it is now seen as ‘morally wrong to have an attainment measure which entrenches low aspirations for children because of their background’ (Department for Education 2010, 68). Whether or not this could have a significant effect on the use of progress-tracking technologies in schools remains to be seen.

References

- Black, P. 1994. Performance Assessment and Accountability: The Experience in England and Wales, *Educational Evaluation and Policy Analysis* 16(2) 191-203.
- Black, P. 1997. Whatever Happened to TGAT. In *Assessment versus Evaluation*. ed. C. Cullingford. London: Cassell.
- Black, P. 1998. Learning, League Tables and National Assessment: Opportunity Lost or Hope Deferred? *Oxford Review of Education* 24(1) 57-68.
- Black, P. and D. Wiliam. 1998a. Assessment and classroom learning. *Assessment in Education* 5(1) 7-73.
- Black, P. and D. Wiliam. 1998b. *Inside the black box: Raising standards through classroom assessment*. London: King's College London School of Education.
- Black, P. and D. Wiliam. 2003. "In Praise of Educational Research": formative assessment. *British Educational Research Journal* 29(5) 623-637.
- Black, P., C. Harrison, C. Lee, B. Marshall, and D. Wiliam. 2003. *Assessment for Learning: Putting it into practice*. Maidenhead: Open University Press.
- Centre for Evaluation and Monitoring. 2011. The CEM Centre. <<http://www.cemcentre.org/>> (accessed 3 February 2011).
- Croll, P. 1990. Norm and Criterion Referenced Assessment: Some reflections in the context of assessment and testing in the National Curriculum. *Redland Papers* 1: 9-12.
- Department for Children Schools and Families. 2008. *Gaining Ground: Improving progress in coasting secondary schools*. Nottingham: DCSF Publications.
- Department for Education. 2010. *The Importance of Teaching: The Schools White Paper 2010*. London: HMSO.
- Department of Education and Science and the Welsh Office. 1988. *National Curriculum: Task Group on Assessment and Testing - A Report*. London: DES.
- Feinstein, L. 2003. Inequality in the Early Cognitive Development of British Children in the 1970 Cohort. *Economica* 70: 73-97.
- Foucault, M. 1975. *Discipline and Punish*. London: Penguin, 1991.
- Foucault, M. 1976a. "Society Must Be Defended". London: Penguin, 2004.
- Foucault, M. 1976b. *The Will to Knowledge*. London: Penguin, 1998.

- Foucault, M. 1978. *Security, Territory, Population*. Basingstoke: Palgrave Macmillan, 2007.
- Gibson, A. and S. Asthana. 1998. Schools, Pupils and Examination Results: contextualising school 'performance'. *British Educational Research Journal* 24(3) 269-282.
- Gipps, C. 1992. National Curriculum Assessment: a research agenda. *British Educational Research Journal* 18(3) 277-286.
- Hacking, I. 1990. *The Taming of Chance*. Cambridge: Cambridge University Press.
- Hall, K., J. Collins, S. Benjamin, M. Nind, and K. Sheehy. 2004. SATurated models of pupildom: assessment and inclusion/exclusion. *British Educational Research Journal* 30(6) 801-817.
- Hoskin, K. 1979. The Examination, Disciplinary Power and Rational Schooling. *History of Education* 8(2) 135-146.
- Kimberley, K., I. Hextall, H. Torrance and B. Moon. 1989. Review Symposium: National Assessment and Testing: the TGAT Report. *British Journal of Sociology of Education* 10(2) 233-251.
- Levacic, R. and P.A. Woods. 2002. Raising School Performance in the League Tables (Part 1): disentangling the effects of social disadvantage. *British Educational Research Journal* 28(2) 207-226.
- Lindsay, G. and A. Lewis. 2003. An Evaluation of the Use of Accredited Baseline Assessment Schemes in England. *British Educational Research Journal* 29(2) 149-167.
- Ofsted. 2009. *Twelve outstanding secondary schools: Excelling against the odds*. London: Ofsted.
- Porter, T. 1986. *The Rise of Statistical Thinking 1820-1900*. Princeton: Princeton University Press.
- Ray, A. 2006. *School Value Added Measures in England: A Paper for the OECD Project on the Development of Value-Added Models in Education Systems*. London: Department for Education and Skills.
- Reay, D. and D. Wiliam. 1999. 'I'll be a nothing': structure, agency and the construction of identity through assessment. *British Educational Research Journal* 25(3) 343-354.
- RM Data Solutions. 2011. RAISEonline – Groundbreaking online education reporting tools. <<http://www.forvus.co.uk/casestudies/dcsfsted-raiseonline/default.aspx>> (accessed 4 February 2011).
- Siegel, H. 2004. High stakes testing, educational aims and ideals, and responsible assessment. *Theory and Research in Education* 2(3) 219-233.
- Strand, S. 2006. Comparing the predictive validity of reasoning tests and national end of Key Stage 2 tests: which tests are the 'best'? *British Educational Research Journal* 32(2) 209-225.
- Thomas, R. M. 2005. *High Stakes Testing: Coping with Collateral Damage*. Mahwah: Lawrence Erlbaum Associates.
- Torrance, H. 1992. Research in Assessment: a response to Caroline Gipps. *British Educational Research Journal* 18(4) 343-349.
- Troman, G. 1989. Testing Tensions: the politics of educational assessment. *British Educational Research Journal* 15(3) 279-295.
- Weeden, P., J. Winter and P. Broadfoot. 2002. *Assessment: What's in it for schools?* Abingdon: RoutledgeFalmer.
- Whetton, C. 2009. A brief history of a testing time: national curriculum assessment in England 1989-2008. *Educational Research* 51(2) 137-159.

- Wiliam, D. 1992. Value-added Attacks: technical issues in reporting national curriculum assessments. *British Educational Research Journal* 18(4) 329-341.
- Wiliam, D. and P. Black. 1996. Meanings and Consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal* 22(5) 537-548.
- Wyse, D. and H. Torrance. 2009 The development and consequences of national curriculum assessment for primary education in England. *Educational Research* 51(2) 213-228.