futurelab
innovation in education

Assessment and Social Justice

a Futurelab literature review: Report 16

John Gardner, Queen's University, Belfast
Bryn Holmes, Concordia University, Montreal
Ruth Leitch, Queen's University, Belfast

KEY TO THEMES OVERLEAF

Key to themes

Futurelab understands that you may have specific areas of interest and so, in order to help you to determine the relevance of each project or publication to you, we have developed a series of themes (illustrated by icons). These themes are not intended to cover every aspect of innovation and education and, as such, you should not base your decision on whether or not to read this publication on the themes alone. The themes that relate to this publication appear on the front cover, overleaf, but a key to all of the current themes that we are using can be found below:

Digital Inclusion – How the design and use of digital technologies can promote educational equality

Innovative Teaching – Innovative practices and resources that enhance learning and teaching

Learning Spaces – Creating transformed physical and virtual environments

Mobile Learning – Learning on the move, with or without handheld technology

Learner Voice – Listening and acting upon the voices of learners

Games and Learning – Using games for learning, with or without gaming technology

Informal Learning – Learning that occurs when, how and where the learner chooses, supported by digital technologies

Learning in Families – Children, parents and the extended family learning with and from one another

For more information on our themes please go to www.futurelab.org.uk/themes

About Futurelab

Futurelab is passionate about transforming the way people learn. Tapping into the huge potential offered by digital and other technologies, we are developing innovative learning resources and practices that support new approaches to education for the 21st century.

Working in partnership with industry, policy and practice, Futurelab:

- incubates new ideas, taking them from the lab to the classroom
- offers hard evidence and practical advice to support the design and use of innovative learning tools
- communicates the latest thinking and practice in educational ICT
- provides the space for experimentation and the exchange of ideas between the creative, technology and education sectors.

A not-for-profit organisation, Futurelab is committed to sharing the lessons learnt from our research and development in order to inform positive change to educational policy and practice.

Photos on pages 23 and 29 courtesy of Inclusive Technology, www.inclusive.co.uk.

CONTENTS

1. EXECUTIVE SUMMARY

Social justice refers to the concept of a society affording individuals and groups fair treatment and an impartial share of the benefits of that society

Social justice refers to the concept of a society affording individuals and groups fair treatment and an impartial share of the benefits of that society. It is crucial in relation to children and young people who have little control over their environment or circumstances, and have little say over much of what happens to them in schools in the name of testing and assessment.

While aspects of educational assessment may contribute to the impacts of social injustice, it must be emphasised that its inequities pale to insignificance in the face of the impact of a whole litany of adverse circumstances that many of our most disadvantaged young people experience. These can include any combinations of sub-standard accommodation, low income, community environments wracked by criminality, violence and substance abuse, poor health and life expectancy, dysfunctional family life, lack of value put on education and, indeed, a poor quality of education provision itself.

Assessment may be categorised as having two main purposes, which in some circumstances may be combined in the same assessment process: assessment to support learning (assessment for formative purposes) and assessment of learning outcomes (assessment for summative purposes).

Disadvantage in assessment contexts may arise from environmental factors (eg poor quality schooling, inadequate access to schooling, poverty) or personal attributes that are treated by others in an unfair fashion (eg language and culture differences, gender).

Several widely accepted codes of practice serve to advise on fair practices in assessment, and these include the AERA Standards for Educational and Psychological Testing and the JCTP Code of Practice (references in following text). Articles 3 and 12 of the United Nations Convention on the Rights of the Child also offer the potential to safeguard students from unfair practice and to give them a voice in assessment processes that may affect their lives. This will be increasingly important in high stakes assessment contexts.

E-assessment currently offers solutions to some types of disadvantage through the use of assistive technologies, but can be improved through the further development and appropriate adoption of such processes as automated marking, universal design, adaptive testing, parallel translation. Handheld technologies provide increasingly flexible means of accessing learning resources and activities. They also provide a means for dynamic engagement with personalised assessments such as coursework, research projects and e-portfolios.

E-assessment also has considerable potential to individualise traditional assessment contexts through adaptive testing, and to provide an increasingly sophisticated menu of assessments that enable a student's full range of learning to be assessed. Ultimately, the prospect exists for current Web 2.0 tools to develop further and support disadvantaged students with personalised learning networks. These will provide low-stakes assessment systems that endeavour to put a premium on learning, motivation and self-esteem. They will also enable students to choose how and when they wish to be assessed instead of restricting them to the assessment provisions laid down by conventional curricula and qualifications frameworks.

2. SCOPE OF REVIEW

This review examines the role that educational assessment plays within a social justice context. It relates mainly to educational assessments for school-age students but draws on examples and experiences across the whole spectrum of ages and phases. Our concept of assessment is all-embracing. It covers assessments that range from the ad hoc classroom interaction between students and teachers at the formative end of the spectrum to high-stakes selection for jobs and university places at the summative end. The pursuit of social justice is viewed as a process that aims to ensure that differences in people's opportunities to access society's benefits are recognised and addressed in order to maximise equity of outcome. Assessment for social justice therefore takes in any form of assessment that has the potential to reduce the effects of a variety of disadvantages that learners may experience.

As a review, this work will focus on the 'now' in assessment. Clearly, however, aspects of this 'now' very much hint at developments on the horizon, almost exclusively in technologically supported assessment. The scope of the review will therefore also involve a degree of speculation about desirable developments for the future.

3. SOCIAL JUSTICE

There is always a well-known solution to every human problem
- neat, plausible and wrong.
(Mencken 1920)

Social justice is a key concept in many educational discussions today. At the most basic level we associate it with practices that assure 'fairness' in a quest towards a type of utopian society. In this society everyone has equal opportunities and access to resources in order to thrive and succeed, despite any disadvantages that may arise as a result of their gender, race, socio-economic status or language. In academic discussions (eg Thrupp and Tomlinson 2005), social justice is considered a complex and multi-faceted concept, which is interpreted differently in differing contexts. However we might interpret it, social justice in its widest sense is crucial in relation to children and young people who have little control over their environment or circumstances. To this we can safely but regrettably add that they also have little say over much of what happens to them in schools in the name of testing and assessment.

In general terms, social justice refers to the concept of a society affording individuals and groups fair treatment and an impartial share of the benefits of that society. Proponents of social justice such as Rawls (1971) and Fraser (1997) have developed different theories or models of the underlying principles and practices of what constitutes social justice. With respect to their application to education and specifically assessment practices, one of the most useful conceptual frameworks is that of Cribb and Gewirtz (2003), which describes three forms of social justice, namely distributive, cultural and associational.

Distributive justice

Distributive justice concerns the ways in which educational opportunities and access are distributed and where resources are redistributed so that the circumstances for the least advantaged are improved. Distributive models of social justice are therefore reflected in compensatory programmes which allocate resources and opportunities specifically for the disadvantaged. Within this approach, the measurement of inequalities in students' attainment is seen as an important tool for tackling inequalities. If a school is deemed to be failing its students, the Office for Standards in Education (Ofsted) can put it under 'Special Measures', which require it to improve its education provision and outcomes (judged primarily by national test results) over a specified time period. During this period additional resources are provided, reflecting the distributive justice approach.

The Department for Children, Schools and Families (DCSF 2008) claims that since 1997 over 1,400 schools were turned around by this additional help while 200 faced the ultimate sanction and were closed. Of the closed schools, 57 were replaced by new schools (the 'Fresh Start' dimension of the policy) and only three of these have proved unsuccessful. Widening access and participation agendas are other parts of the government's top-down approach to improve educational opportunities for the poor, marginalised or excluded. Atweh (2007) has argued, however, that such distributive justice can create a superficial equality of opportunity if it does not question the roles of curriculum, pedagogy and assessment in creating educational inequality. Similarly, it may prove ineffective if it does not take into account the historical and political reasons that underlie the social, economic and racial divisions and inequalities that feed disadvantage.

Cultural justice

Cultural justice promotes recognition of cultural difference with respect and tolerance for differing cultures and communities. Cribb and Gewirtz (2003) view this as one means of removing barriers to social justice, while Demie (2005) exemplifies the concept with evidence from 13 case study schools in Lambeth, London where the performance of Black Caribbean students has 'bucked the trend'. Their results were shown to be demonstrably above the national average at Key Stage 2 and GCSE. This is explained not simply as a result of using performance data but also through promoting school cultures of high expectations, an inclusive curriculum and community participation. Along with their commitment to equal opportunities and a clear stand on racism, the schools tangibly displayed the key conditions and practices within a cultural justice model. Such examples of good practice (see also Maguire, Wooldridge and Pratt-Adams 2006) may demonstrate the positive impact of culturally just practices, but the pursuit of cultural justice also has the potential to be unjust. For example, Taylor and Gunter (2008) argue that it may not enable individuals from the cultures involved to contribute towards fundamental decisions that influence them and it may inadvertently accentuate the socio-cultural divisions within education and society.

Associational justice

Associational (or participatory) justice is enjoyed where those who are not part of the dominant group contribute to the decision-making process that structures their experiences and opportunities. Arguably, in this context, social justice is based on the concepts of human rights and entitlements, and extends to children's rights through the implementation of the United Nations Convention on the Rights of the Child (UNCRC 1989). In educational contexts, children are generally not afforded an opportunity to participate in decision making on matters of significance to their educational experiences. Smith (2006) argues that research that focuses on assessment and testing and the implementation of the rights of the child is relatively rare, but some work is available. While not explicitly focusing on rights, Duffield et al (2000) have certainly looked at students' voice in relation to how they and their schools perceive achievement. Leitch et al (2007) have also demonstrated that children can be consulted directly and inclusively by policy makers on matters of educational significance such as assessment policy and practice. The children's views on how their performance should be recorded in 'annual pupil profiles' influenced the final format of the assessment chosen by government and is a clear illustration of associational justice in practice. Mostly, however, students can be impeded from full participation in assessment design by denying them a voice, much less an equal voice, in the development of assessment policy and practice.

Any pursuit of social justice in education, then, can be theorised to require close attention to the distribution of educational resources (eg effective and well-resourced schools, and good teachers), recognition and respect for cultural differences (and the removal of barriers relating to these) and participation (not just inclusion but active participation in designing and effecting the education provided). There are other dimensions (see for example Lynch and Baker 2005, who include equality of "love, care and solidarity"), but these three form the basic framework for the field.

Before moving on to assessment, and later its influence in social justice contexts, it is worth making our position clear on the part that educational assessment plays in the social justice agenda. Underlying the models above are forces that greatly outweigh the role of any form of assessment in contributing to social injustice. Kozol's (1991) observations on the "savage inequalities" of urban living are no less relevant today than they were in 1991. So too is Gordon's 1995 perception (p360) that "...equity problems in assessment are largely secondary to the failure to achieve equity through educational treatments". And Reay (2006) sees the "...zombie stalking English schools: social class and educational inequality" (p288) as being largely unchanged in 50 years. Thrupp and Lupton (2006) question policies that seem to assume consistency in school contexts when in fact a one-size-fits-all approach is manifestly unsuccessful. A stark account of the different outcomes typical of different types of schools has also been published by Levacic ̆and Marsh (2007). Their review of the examination results from schools in the 20 English local authorities that are still either wholly selective (10) or partly selective (10) showed that "...the effect of attending a [non-selective] secondary modern school was on average 1 grade less at GCSE compared to a [non-selective] comprehensive school" (p163) and five grades less than students in (selective) grammar schools. Taking account of funding and costs, they conclude that students:

> ...are academically disadvantaged by attending secondary modern schools, which in selective local authorities do not receive sufficient additional funding to offset the depressing effects on attainment of the increased social segregation arising from a selective system.
> (Levačić and Marsh 2007, p155)

We must acknowledge that inequities in assessment pale to insignificance in the face of the impact of a whole litany of adverse circumstances that many of our most disadvantaged young people experience. Take your pick from sub-standard accommodation, low income, community environments wracked by criminality, violence and substance abuse, poor health and life expectancy, dysfunctional family life, lack of value put on education and, indeed, a poor quality of education provision itself.

4. ASSESSMENT

The whole subject of examinations looms very large in the vision of the public and is apt to be seen out of its true proportions, mainly because it is the one portion of school business which is recorded in newspapers. We shall perhaps arrive at right notions about it more readily, if we first consider the business of examinations wholly subordinate to that of education, as part of the work of a school.
(Fitch 1898, pp158/59)

Assessment purposes

Ultimately assessment has to address a variety of learning outcomes. These have traditionally fallen into academic or vocational categories, but today's 21st century or 'new' learning skills are making demands for 'new' assessments. The new learning that is often spoken about is notoriously difficult to pin down but QCA (2008a) has identified some of the skills involved as team working, independent enquiry, self-management, reflective learning, effective participation and creative thinking. Baker (2007) has also argued the need to assess "...adaptive problem-solving, risk assessment, managing distraction, self-management and changeable roles". Meeting the challenges of assessing such complex learning outcomes will take considerable ingenuity and innovation.

Perusing the large majority of the assessment literature will likely lead to the conclusion that there are two types: formative and summative. However, we would argue that these are purposes to which assessment is put. It is not the assessment (judgement or grade; technique or test) that is formative or summative; it is how it is used. Of course, some modes of assessment are better suited to one or the other. The two main purposes of assessment, then, are to act in support of students' learning (usually termed formative assessment or assessment for learning), and to report the students' levels of attainment (usually called summative assessment). On this basis, therefore, we will use the terms formative assessment and summative assessment on the understanding that they do not describe the assessment per se; they reflect the uses to which the assessment is put.

Formative assessment in UK schools is an activity in which teachers interact continuously with their students, assessing where they are in their learning and providing guidance on the next steps to improve their learning. This form of assessment, when carried out constructively, is beneficial to all types of students but comes into its own in supporting students who for whatever reason (eg second language, cultural or personal isolation) are withdrawn, disaffected, have low self-confidence or low academic self-esteem. In the US and in higher and further education in the UK, a more structured and substantially different conception of formative assessment is often encountered. This is the use of frequent short tests, usually multiple choice, to identify (diagnose) students' deficits in knowledge and understanding. In this form, it feeds more directly into the next steps in teaching rather than learning and some see it as a form of assessment that can be used formatively to support personalised teaching or summatively in profile or aggregate form to report progress.

Continuous assessment, in the sense of project work or portfolio-based assessment rather than frequent testing, can also provide considerable support for the learning process. Here the ongoing subjective assessment of their teachers can build students' confidence and a sense of achievement over time. There can be controversy when such assessments are used for national level assessment, with concerns periodically expressed about the teachers being inappropriately supportive (ie giving substantive help – see, for example, Mansell's 'Mrs Ford', 2007, p67) or plagiarism.

There is no particular concern for the assessment to support learning and the range of its use remains narrowly functional

Generally cast as discrete events, assessments designed exclusively for summative purposes may offer little or no feedback to the students on their strengths and weaknesses or on how their performance could be improved. It almost goes without saying, therefore, that existing examinations and tests would struggle to meet the reasonable aspirations of many educational experts, who consider that:

_ assessment of any kind should ultimately improve learning

_ assessment methods should enable progress in all important learning goals to be facilitated and reported

_ assessment methods should promote the active engagement of students in their learning and its assessment

_ assessment should enable and motivate students to show what they can do (Gardner et al 2008, p16).

For many the raison d'etre of testing is simply to give an 'objective' assessment of a student's performance in a particular domain of learning. The reliability of such assessments is considered to be strengthened by the objectivity of assessors who do not know the students or the schools involved. There is no particular concern for the assessment to support learning and the range of its use remains narrowly functional.

Yet there is no academic or assessment-related reason why examinations and tests should not be used formatively. In fact it can be very beneficial, as argued by Harlen (2006) and amply demonstrated by Black et al (2003) and Carter (1997). Carter allowed her students to take greater charge of their learning by returning their examination scripts. She confirmed that there were errors in the scripts but did not identify them. Instead she encouraged the students to find and correct them collaboratively. Could this be done with existing National Curriculum tests, GCSEs, A-levels, Highers etc? Theoretically yes, but practicably no, or at least not without a major organisational change that would provide a period of feedback and reflection as part of the examination cycle, after the results are released.

Most examinations and tests are designed primarily to assess whether a student has made a particular grade or to identify what grade they have achieved, according to some standardised or tailored scale Their primary purpose, therefore, is not to support learning but to categorise the learner – arguably to identify what they cannot do rather than what they can do. We recognise that it is legitimate for assessment to have a variety of purposes, for example to contribute to assessing school performance (accountability), enabling entry to a university (selection) or to enable the award of qualifications (the US 'credentialling'). However, we see it as being at least as important to ensure these relatively stressful events have a positive impact in students' lives through supporting improvements in their learning, motivation and self-esteem. Indeed, there are strongly held views that in these respects national and standardised testing can have significantly detrimental effects (see, for example: ARG 2002; Johnston and McClune 2000; Leonard and Davey 2001; Remedios et al 2005; and Putwain 2007). If we accept that some forms of assessment, notably tests and examinations, may have negative effects on students, then it is not unreasonable to suggest that such effects may be magnified for those who enter the process already labouring under disadvantage.

As will be outlined later, there is a range of adjustments made in existing tests and examination processes in an attempt to improve access to assessments for those for whom it would normally be difficult. Trialling and consultation are also widely used to reduce such problems as gender bias and cultural unfamiliarity in the content of

such assessments. Clearly assessment methods in the future should continue to be designed and administered to address the effects of any disadvantages. But can we expect them to do better? Can we expect them to shift emphasis from showing what a student cannot do to what they can do? Can we ensure that all tests support learning by motivating and promoting the engagement of the students in their learning? These are tall orders for current methods of assessment. Will the future be any different?

Assessment types

There are many ways to categorise types of assessment, but for our purposes we will use the following groupings: national tests, standardised tests and extended tasks.

National tests

These may be compulsory (eg the recently discontinued Key Stage 3 National Curriculum Assessments in England, also known as SATs - Standard Assessment Tasks) or voluntary (in the sense of selecting GCSE subjects, for example). In the main, they are medium to high-stakes, defined as having outcomes that influence life choices. Medium-stakes tests may operate at the school level, for example influencing choices of subjects in subsequent school years, while high-stakes tests may act as gatekeepers to over-subscribed programmes in third-level education. They are generally time-bound and may comprise any one or combinations of multiple choice, structured questions and short essay-type questions. 'Objective' reliability is privileged over validity and the students' answers are scored against prescribed mark-schemes by external assessors – a form of fairness that is based on everyone being marked in the same way. National tests such as the English SATs are generally weak on validity, for example being unable to assess more than a small proportion of the curriculum under study.

Standardised tests

In some circumstances, certainly in countries such as the US where psychometric testing still holds sway, national tests may also involve standardising the tests. An element of the design of these tests is that their scores are compared with representative samples of the target student group. Students who take the tests are then ranked according to the range of performances identified in the normative standardisation of the target group. In the UK various such tests are used diagnostically for cognitive ability, reading ability and so on. For example, the Qualifications and Curriculum Authority (QCA 2008b) offers a list of diagnostic tests (which they are careful not to endorse!) for teachers to use to help them decide if students need special support. All standardised tests purport to offer 'objective measures' of such complex attributes as IQ, attitudes, dispositions, reading skills (eg word recognition and spellings) and writing speed, but we are reminded of Alsthuler and Schmautz's (2006, p11) comment: "Standardised assessment of diverse peoples is, arguably, an oxymoron".

Extended tasks

Students may be assessed on a significantly large piece of work either on a continuous basis during its creation or at the end of a relatively long period of developing it. The assessment vehicles for these extended tasks include research and project reports, essays and portfolios. Unlike the normative basis for the tests above, the quality and level of achievement in these types of assessment are generally set against relatively broad criteria that can allow the student a considerable degree of expression. The assessment itself is usually based on the subjective judgment of the assessors, who in many cases are the students' teachers. Much of the professional endorsement for these types of assessment derives from their capacity to engage the students in 'authentic' activities, that is, activities that bear some relationship to life experiences or the practical expression of the concepts involved.

The growth in computer-based testing in recent years is quite phenomenal

Moving on from pen and paper

Aside from extended tasks, which generally only require a rubric to guide students on topic choice and ways of working, the majority of assessments can be delivered by computer. That is, they can be reproduced on screen as they appear on paper (sometimes called "paper behind glass", Becta 2006, p3). One major difference from paper versions, of course, is that the students complete their responses on the computer also.

Becta (2006, p3) posits three main benefits of using computers to support assessment:

_ Improving traditional processes by, for example, scanning exam scripts for electronic forwarding to markers, online marking, electronic delivery of exam scripts for printing at exam centres and the online completion of 'traditional' exam scripts.

_ Extending the limits of traditional practice by, for example, online tests which extend what can be assessed through the use of multimedia, simulations and 'drag and drop' mechanisms; and making available item banks and randomised question choices to provide assessment on-demand.

_ Technology in the service of learning that uses e-assessment to provide ongoing formative assessment (with the integration of assessment and learning content), progress tracking, goal setting, feedback to the learner and practitioner, diagnoses of understanding and levels of ability, recording of achievement and storing of evidence of varying types (for example audio and video files of practical work).

At this point we turn our attention to e-assessment itself.

5. E-ASSESSMENT

The capacity and technology for e-assessment is developing rapidly. We recognise the potential of this development to enhance the quality and efficiency of assessment. E-assessment should not be construed as limited to quick multiple choice testing; it has the potential to test learners in both structured and unstructured environments... (Tomlinson Report 2004, para 156, p63)

In this manner the Tomlinson Report gave its backing to embedding e-assessment in future educational developments. However, definitions of e-assessment vary according to perspective. For example, the one favoured by the Joint Information Systems Committee (JISC) (with its orientation to third-level education) and the Qualifications and Curriculum Authority (QCA) is presented as follows:

e-Assessment is the end-to-end electronic assessment processes where ICT is used for the presentation of assessment activity, and the recording of responses. This includes the end-to-end assessment process from the perspective of learners, tutors, learning establishments, awarding bodies and regulators, and the general public. (JISC/QCA 2007, p6)

A more educationally oriented definition is offered by Becta (2006, p2):

The (electronic) process by which learner progress and understanding is assessed. This can be:

- diagnostic (to assess current levels of knowledge and understanding in order to target future learning appropriately)
- formative (to support and feed back into current learning)
- summative (to assess knowledge and understanding at the end of an episode of learning, usually equated with a formal award).

These are up-to-date statements of what e-assessment is and can do, but the pursuit of machine-based ways of assessing student work can be traced as far back as Pressey (1926) who launched a 'teaching machine' that was a "...simple apparatus which gives tests and scores". It was of course a very rudimentary machine and subsequent developments remained so until the arrival of digital technologies brought the prospect of major advances in the late 1970s. The three decades since then have seen many developments that fall largely into two categories: improvements or efficiencies in test administration, and adaptive testing. The former is primarily the delivery of traditional testing online or at least on a computer, and most examination bodies have espoused the many administrative benefits such developments offer. In 2004, the QCA outlined a blueprint for e-assessment that aimed to have at least two on-demand, online GCSE subject examinations by 2008 (QCA 2004). In light of the recent marking debacle over England's National Curriculum tests (BBC 2008a), QCA has also expressed the hope that by 2010 on-screen marking will solve the problems that have been experienced in distributing scripts and recording marks (Boston 2008).

The administrative benefits arising from the use of computer-based approaches include the digital delivery of tests 'anywhere, anytime', personalised delivery, direct links to digital learning materials and courses, automated marking and immediate feedback. The growth in computer-based testing in recent years is quite phenomenal. For example, GL Assessment claims that it has had 10,000-plus online assessments undertaken per day in the peak months of September and October (GL 2008) and that the results for the largest single sitting (3,000 candidates) of its standardised Cognitive Abilities Test were returned within five minutes. However, the benefits can be problematic.

The ability to make the work available electronically, with a varied use of media such as text, video and audio, gives the e-portfolio a boost over traditional paper or artefact-based collections

For example, there are considerable security challenges to be surmounted if the tests are to contribute to high-stakes decision making or if they have to be offered on the continuous basis necessary to fulfil the demands of 'anywhere, anytime'. The costs can be prohibitive for all but the largest assessment agencies if the full 'bells and whistles' facilities of computer-based assessment are to be used in questions and tasks. Boyle (2005, p9) calls these 'sophisticated tasks' with certain core features including: "...media-rich stimulus material (whether graphical, sound, video or animation)" and the test taker is "...required to interact with the stimulus material in a variety of ways".

E-portfolios

One of the most promising forms of e-assessment is the e-portfolio, a vehicle for the types of assessment falling under the extended tasks category above. Becta (2006) has identified five types of usage for portfolios:

_ showcase – enabling students to present and share their work

_ learning and development – setting out the student's learning plans and outcomes

_ lifelong learning summary – a record of the student's achievements over time

_ transaction – a storage system for personal files and utilities

_ evidence – a collection of work to be submitted for assessment.

A comprehensive listing of the desirable features for e-portfolios is offered in a joint QCA/CCEA/SQA/DCELLS document 'E-Assessment: Guide to Effective Practice' (QCA 2007, Annex 6) but in essence the generic features can be summarised as being student driven: the student should choose the content and indicate how it addresses the assessment criteria. Very often the students are also asked to provide their own evaluation of the work they have completed.

Mason et al (2004) offer a simple classification of e-portfolios as being for development, presentation or assessment purposes. This last usage continues to attract a lot of attention, particularly in a wide variety of creative and design-based learning contexts where portfolios have a long-established role in the assessment of work created by students in subjects as diverse as art and engineering design. The ability to make the work available electronically, with a varied use of media such as text, video and audio, gives the e-portfolio a boost over traditional paper or artefact-based collections. However, concerns exist over the assessment time and cost of assessing what can be very complex collections of work. In addition, Twining et al (2006) set out arguments that suggest "...their development and uptake are still at a very early stage" (p57).

Two recent projects have shown how the potential of e-portfolios may be realised. The e-Scape project (Goldsmiths 2007) has successfully developed and trialled a portfolio system for assessing design innovation. Described as a dynamic portfolio process (created over a pre-set six-hour period by the students), it acts to record the students' work directly to a dedicated and secure web server from a variety of input devices including PDAs, speech tools, cameras and digital pens. The portfolio can then be electronically distributed to a number of judges and graded on a comparative pairs basis, ie each portfolio is located on a ranked list according to at least 16 one-by-one comparisons with other portfolios. This establishes a ranked position for each portfolio at which the portfolios below are considered to be less good and the ones above are considered to be better. The benefits of the e-portfolio approach are argued to include enabling the teacher to control task sequencing while an evidence trail is built up automatically in each portfolio on the web server. Real-time learning in practical contexts (the studio or workshop) is supported and can be extended to other performance fields.

Another innovative use of e-portfolios, e-Viva (McGuire 2005), also makes use of handheld devices (in this case mobile phones) but is quite different in that its purpose is to support teachers' judgments through being the stimulus for discussion between them and the students. Students create an e-portfolio containing annotated evidence of their ICT learning 'milestones'. As the name suggests, they are then "...expected to participate in a telephone viva in which they answer a number of pre-selected questions about their working processes and their learning journey" (p267). The combination of viva and an annotation tool for the portfolio is argued to increase opportunities for dialogue between students and teachers and for students to demonstrate the thinking behind their work.

Computerised adaptive tests

A computerised adaptive test (CAT) is quite distinct from a test that is merely presented on computer. Its central design is based on using a bank of items for which the specific levels of difficulty have previously been established. Each time students are presented with a question their answers are assessed as correct or incorrect. Put simply, the computer then chooses the next question from a range of items of a higher, similar or lower level of difficulty than the one just completed. In this manner, if the students answer correctly they are offered an item of slightly higher difficulty, or indeed similar difficulty to consolidate the measured level of their achievement. If they answer incorrectly, the next item is less difficult or of similar difficulty for consolidation. This analysis of the 'right and wrong' answers then continues to define the choice of subsequent questions as the system 'adapts' to the performance level of the student.

Most CATs do not allow students to review and revise their answers because the sequence of questions is chosen on the basis of previous answers. If the facility did exist, and the students were able to go back and correct 'wrong' answers, the sequence of subsequent item choices and level of difficulty achieved might no longer be valid or reliable. That said, Lilley and Barker (2004) have explored the effect of allowing revision of answers. They have shown that the impact on performance is minimal but benefits may include a reduction in the students' anxiety levels and an increase in their confidence.

CATs have been around for some 50 years and for much of that time they have been quite basic. Renewed interest in adaptive tests burgeoned in the mid-1990s, but this has waned significantly in the face of the many technical and educational issues that surround their use (for example see Way et al 2005 and Stocking et al 2000). Indeed as a result of such problems the world-leading testing agency, the US-based Education and Testing Service (ETS), proposed to abandon CAT versions of its flagship Graduate Record Examination (GRE) tests in 2007 and move to a linear testing format which "...is more test taker friendly [than the current computerised adaptive test]... is more familiar to test takers... allows test takers to review, omit, go back and change an answer... [and]... allows test takers to better allocate their time according to their personal test-taking approach" (ETS 2006). These plans were in turn shelved owing to web access problems (IHE 2007).

CATs work best with fixed-response question formats such as multiple choice and can be powerful tools for diagnosing where students are experiencing difficulties in their learning (eg see ALTA, in Ripley 2007). But CATs are generally not used for more complex assessment formats such as essays, though advances are being made in automated essay marking. For example, Nichols (2005) has shown that there is relatively strong agreement between marks awarded for students' writing tasks by expert judges and a system called the Intelligent Essay Assessor, developed by Pearson Knowledge Technologies. Monaghan and Bridgeman (2005) have also shown significant agreement on essay marking between human judges and ETS's e-Rater, sufficient to enable the e-Rater score to contribute to decisions about whether a third marker is required when two judges disagree. Attali and Powers (2008) have explored the prospects of developing a scale to assess writing using a modified selection of e-Rater measurement targets, namely: grammar, usage (explained as errors such as wrongly used words, preposition errors etc), mechanics (eg spelling and punctuation errors), style, essay length, vocabulary and word length. They too identified high correlations between human judges and the e-Rater system but acknowledged that the writing tasks cannot address the full gamut of writing skills. Despite the existing technical and academic challenges, though, it is likely that CATs will continue to improve and attract attention in assessment contexts.

6. THROUGH NO FAULT OF THEIR OWN

If a man does not keep pace with his companions, perhaps it is because he hears a different drummer. Let him step to the music he hears, however measured or far away.
(Thoreau 1854, p343)

At the heart of the problem is the fact that most assessments are designed for students who fit a societal norm

Through no fault of their own, assessment can serve some students badly. It can happen because the design and content of the assessment instruments discriminate against them in some way or the assessment process itself is unfair. However, it is worth reiterating that the most dominant factor underpinning disadvantage and differentiated outcomes, for many students and their communities, is the social context in which they live. Assessment is often the process that reveals disadvantage or discrimination.

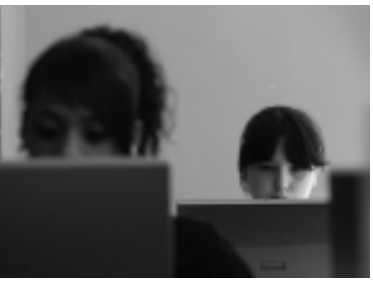
Assessment processes, especially so-called standardised tests, continually produce results that show up gaps in achievement between groups in society. Explanations for the gaps often suggest causes that include the content being biased against the underperforming group's cultural norms, restricted familiarity with majority norms, their higher levels of disaffection and demotivation, or their struggle for equitable resources and learning opportunities in the face of almost overwhelming social and community problems. Not surprisingly, they usually differ from the majority White and English-speaking group in one or more of the following: minority ethnic status, English as an additional language, religion and culture. Other groups, drawn from across the majority and minority communities, suffer from inequitable access to assessment and, perhaps more importantly, the learning opportunities that should underpin their educational development. These groups include students who are disabled or who are isolated from mainstream provision through distance or significant ill-health. At the heart of the problem is the fact that most assessments are designed for students who fit a societal norm – able-bodied, in command of the language of the assessors, appropriately prepared (teaching etc) and full-facultied in terms of capacity to read, write and understand what is presented to them. In addition to group effects, individual students may also suffer directly from assessors' conscious or unconscious negative dispositions to normal student attributes such as gender and ethnicity.

Fairness has troubled educational assessors for decades. Well, most of them anyway. We must never discount the persistent rump of disingenuous educationalists and assessors who consider that the primary role of assessment is to separate the 'chaff from the wheat', the weak from the able, the malingerers from the stalwarts of tomorrow's society. For them, assessment can categorise and organise; it is simply inevitable that those who have will get more, and those who have not – tough luck. As we have indicated earlier, we see assessment as being a process that should be used primarily in support of learning. In all that assessment does it must be fair and seen to be fair, as expressed by the AERA Standards (AERA et al 1999). Camara and Lane (2006) express fairness as a lack of bias in items and tests, examinees having comparable opportunities to demonstrate their standing on a construct, there being comparable outcomes across groups, and individuals and groups having adequate opportunities to learn.

Social justice aims to create circumstances in which society's current inequities are eradicated, or at least their worst disadvantages are mitigated. If a high-stakes assessment is being used, for entry to university for example, social justice demands that students should not be disadvantaged by circumstances and influences that have nothing to do with the constructs under examination. Plummer (2003) would see these circumstances as the 'Big 8' dimensions of diversity: race, gender, ethnicity/nationality, organisational role, age, sexual orientation, mental/physical ability and religion. At the risk of discriminating against some of these, we will look briefly at disability; the combination of language, culture, race, ethnicity and poverty; and gender.

Disability

Since the mid-1990s, the difficulties faced by disabled people have been progressively addressed by robust legislation, most notably the Disability Discrimination Act (DDA 1995, 2005) and its extension, the Special Educational Needs and Disability Act (SENDA 2001).



In the UK, concern for equity is high on every agenda, resulting in considerable research on distributive, cultural and associational justice opportunities and barriers

Under the law, educational institutions cannot treat disabled students less favourably than others in assessment contexts simply because they have a disability (SENDA 2001 for England and Wales, with overlap in Scottish law, and SENDO 2005 for Northern Ireland). 'Reasonable adjustments' have to be made for disabled students. These adjustments, in an assessment sense, are often called modifications or accommodations, though in some quarters these terms are considered to be oppressive language given their implications of a deficit on the part of disabled students or their being a burden on society. Pitoniak and Royer (2001) offer a US perspective on accommodations in a review of the complex psychometric, legal and social policy issues involved, while Cumming and Dickson (2007) offer an analysis, with Australian examples, of the legal implications of not attending to students' special needs. In the UK, the language of discourse is more oriented to ensuring access. For example, in relation to Key Stage 2 National Curriculum Assessment, the QCA says that its tests:

...have been designed to ensure the vast majority of pupils working at the level of the tests can access them. A small number of pupils may require additional arrangements to access the tests.

With the caveat:

The access arrangements must never provide an unfair advantage, the support given must not change the test questions and the answers must be the pupil's own.
(QCA 2008c, p58)

Depending on specific circumstances, the access arrangements may include:

_ additional time: for example, for students with slow reading speeds, including slow information processing, and low writing speeds

_ rest breaks: for example, for students who suffer from fatigue or who have concentration problems

_ readers: for example, for students who normally have material read to them in class; generally the students will have visual impairments or reading ages significantly lower than their actual age

_ scribes: for example, for students who cannot write answers on their own owing to a disability or injury

_ transcripts: for example, for students whose handwriting is likely to challenge the assessor's ability to read it

_ word processors: for example, for students who would normally use a word processor in their class work

_ large print and braille: for example, for students with visual impairments

_ signers and communicators: for example, for students with hearing impairments

_ prompters: for example, for students who have attention problems and need prompting to stay on task.
(Source: QCA 2008c).

There are also options for assistive technology to be used. These include a variety of supports for visually impaired students (eg screen magnifiers, screen readers) and haptic devices for those with physical difficulties (eg various digital input devices - see Techdis 2008 for a useful list of these). However, even when the necessary assistive technologies are available, some disabled students can still experience poor service. A recent example (see the following box) refers to Sam Latif, a blind IT project manager.

Sam, in common with many vision-impaired people, uses text-to-speech conversion software (in this case JAWS) to assist her in her work. In 2004 she studied for a project management qualification and naturally expected to take the final computerised examinations using her assistive technologies. The US-based Project Management Institute refused to allow her to use her own JAWS-enabled computer at the UK test centre and then refused to allow her to install JAWS on a test centre machine. She had to have the exam screens read to her and ended up completing the exam in a much extended eight-hour period. She described the experience as "...really painful and quite upsetting". Sam was perhaps exceptional. Other disadvantaged students may have given up but she persevered with the circumstances dictated by the examination body. Better still, she passed and subsequently won a landmark case against the US company for discriminating against a blind person. Assessment with social justice awareness was not apparent in this case. (Source: OUT-LAW 2007)

Language, culture, race, ethnicity and poverty

We have chosen to consider these issues together because they are often interlinked and interdependent. There is a great deal of research into the types of problems that specific minority groups can experience, and just as in the case of disabled people, these problems are not of their own making. Many of them arise from the ignorance or prejudice of the majority group, whose members in the UK tend to be White, have English as their native tongue and hold considerable sway over central and local governance arrangements. Non-White, English as an additional language (EAL) groups are likely to experience unfairness in educational and employment situations, though White minority groups can also suffer, for example Travellers and non-English speaking migrant workers from Europe. In the UK, concern for equity is high on every agenda, resulting in considerable research on distributive, cultural and associational justice opportunities and barriers, but little of this is in the field of research in assessment.

However, a major UK survey of over 15,000 Year 9 students and their parents (Strand 2008) has shown that around three points separate the higher levels of White students' performances at Key Stage 3 from those for Pakistani, Black Caribbean and Black African students in the sample. Strand equates this to a whole year of progress in terms of National Curriculum levels. However, the deficit is much less than the ten points between students from higher managerial and professional families and those with unemployed parents or guardians, or the nine points between students with mothers educated to degree level and those with mothers who had no educational qualifications. Such results point to the huge impact that poverty-related circumstances can have on educational attainment, and the potential for educational practices and provision (and by association, assessment practices) to reproduce social inequalities. Demie and Strand (2006) have shown that language, specifically the stage of English fluency, is a major factor in performance at GCSE. Others have questioned the ability of teachers to ensure an equitable learning environment. For example, Walters (2007), arising from a small-scale case study of Year 3 Bangladeshi children, argues that "...despite policy and practitioner discourses of multiculturalism and diversity... [the three] pupils were supported to assimilate, to fit in and to be like other pupils rather than supported in their language development and learning" (p99). This lack of cultural justice is not uncommon and Kelly and Brandes (2008) provide guidance on how teachers should develop socially just classroom practices by fostering students' self-development and self-determination.

Criticism of schools and teachers in dealing with under-performance reaches serious levels when it comes to government perspectives and policies in England, though according to Torrance (2003, p905) "...with rather less visceral hatred of the teaching profession [in Scotland, Northern Ireland and Wales] than has been manifested by central government in London".

The notion that schools are failing students and that increased testing is needed to raise standards certainly pervades English government policy. By way of contrast, Hayward (2007) argues that Scotland has adopted an holistic approach (curriculum, pedagogy and assessment reform) to achieving greater social justice. Undoubtedly there are schools and teachers failing to provide an adequate education to their students, but the proposition that testing and increased accountability raises standards of education across the board is a very dubious theory, if not simply a non-sequitur. In educational assessment contexts, much of the research on the effects of testing comes from the US. Its conclusions generally point to the impact of educational and social contexts in causing and sustaining the gaps in achievement between groups. Slavin and Madden (2006), for example, state that "African Americans, on average, attend schools that are far less well funded than those attended by Whites; their teachers are less highly qualified, and their families are more likely to suffer from the ills of poverty, which have a direct bearing on children's success in schools"(p390).

Perhaps the most researched approach to using assessment as a means to address inequities is the US No Child Left Behind Act (NCLB 2002). It is a hugely complex policy and we cannot do it justice here. In the educational world it is largely maligned, sometimes emotionally – sometimes with scientific authority. An example of the former is Keyes's (2007) claim that NCLB is widening the gap between poor and minority students and wealthy White students by restricting the former to answering simple questions in basic skills education while the latter learn to ask complex questions. Nichols et al's work (2005) is an example of thorough research that casts doubt on the efficacy of NCLB in meeting its stated objectives. They suggest that the problems caused by the high-stakes, standardised tests, which are used to inform actions under NCLB, will actually affect minority students disproportionately.

Fusarelli (2004) takes the view that there is much to commend in NCLB, but when asking if the provisions of NCLB will be enough to compensate for the effects of poverty, which fall disproportionately on 25% of African American, Hispanic and Native American children, his answer is two-fold. 'Yes' is the simple answer if the assumptions in NCLB hold, but:

> ...the complex response may be that the intent is laudable, the prescriptions inadequate, the execution poor, and ultimately the results insufficient to narrow the achievement gap and truly leave no children behind. (p89)

By way of building his argument he cites the extent to which the withdrawal of Title 1 funding for schools whose students fare badly – most frequently they are "...impoverished schools filled with students of colour" (p87) – is a sanction that defies logic. The states whose schools do best under NCLB are those that are predominantly White, for example Minnesota (89%), Wyoming (92%) and Connecticut (82%). In a worst case scenario, he argues, the results of the standardised testing approach may include the lowering of state standards, and "...the cure will be worse than the disease" (p87). In a similar vein, this time in England, Levačić and Woods (2002) and Levačić and Marsh (2007) have argued that schools with high proportions of disadvantaged students may suffer from reducing budgets over time as the proportion of disadvantaged students tends to increase and their position in the local hierarchy of schools in turn deteriorates. In such circumstances social injustice is likely to be deepened as performance on national assessments continues on a downward trajectory.

However, assessment cannot lay all the blame for poor performance at the door of social circumstances. Altshuler and Schmautz (2006) point to assessment design and content not recognising cultural differences as partly explaining the large differences in performance on state tests between White and Hispanic students. They argue that White, middle-class values such as promoting independence and self-expression are inherent in test designs and that they systematically discriminate against Hispanic students by conflicting with their cultural values of interdependence.

The experience in the UK and US of using assessment, specifically testing, to promote actions that ameliorate and ultimately remove inequities in achievement for students from poor and minority groups, leaves open the question as to whether it is or can be effective in the face of the major underlying ills that afflict such disadvantaged communities. Technologically supported education provision may, however, be one way forward. For example, the Department for Children Schools and Families' £90 million Computers for Pupils initiative (Becta 2008; Teachernet 2008) is putting e-learning and e-assessment into the homes of pupils who experience serious disadvantage. Perhaps such digital equity initiatives will go some way to ameliorating the effects of disrupted or ineffective schooling.

Gender

There has been considerable debate over many years about the differential performance of males and females in a variety of assessment contexts. In some instances females as a group do better than males (eg in assessments requiring extended writing such as coursework assessment, though the effect is challenged by Elwood 2005 and Woodfield et al 2005) and in others it is males as a group that do better (eg in multiple choice tests, though item content may be an important determinant). Various generalised differences in gender dispositions (for the purposes of this review we choose not to differentiate between biological sex and socially constructed gender) such as risk-taking or willingness to please are offered to explain differences in performance, while strategies to reduce the gap in specific subjects (eg girls performing less well in physics than boys, boys doing less well in English than girls) are often addressed from item design and content perspectives. In one example (McCullough 2004), items with a perceived male orientation in a physics test (eg cannonballs and rockets) were replaced by items with a perceived female orientation (eg stuffed animals, jewellery) and the achievement gap was reduced. However, as might be expected, the results showed that this had been achieved primarily by reducing the male scores.

This item-level approach is the basis of the many DIF studies (differential item functioning) that identify items that favour one sub-group over another, eg boys over girls. Once identified, they can be 'neutralised' or removed, making the test that little bit more fair. But addressing the 'obvious' may not be the answer. Even the recent major revisions in the design and content of the US SATs have been described, on the basis of the College Board's own evaluations, as failing to address the former version's under-prediction of first year college performance of women, minority groups and students whose best language is not English (FairTest 2008).

According to a number of commentators, the gap between male and female performance in a variety of educational contexts is relatively narrow. For example, Strand's (2008) survey of 15,000 students' performance at Key Stage 3 revealed a very small differential of 0.4 points (less than two months' progress according to his scale of three points equating to one year). Gray et al (2004) reported 2001 figures that showed the proportion of girls achieving five A*-C grades at GCSE was higher than that for boys, at 54% compared to 43%. However, on the basis of deeper analysis, which revealed girls' and boys' performance levels were very similar in half of all schools, they argue that school effects are very important factors. The most recent figures from the Department for Education and Skills (DFES 2007) show that the proportion of girls achieving five A*-C grades at GCSE in 2006 was still higher than that for boys (63% compared to 54%) but with a slight decrease in the gap.

There are other indications that gender performance gaps may be decreasing. Recent work carried out by the American Association of University Women, AAUW (reported by Strauss 2008), concludes that gaps between boys and girls in the National Assessment of Educational Progress have narrowed or at least stayed the same in various assessment contexts over several decades. A co-author of the AAUW report is quoted as saying "If there is a crisis, it is with African American and Hispanic students and low-income students, girls and boys" (Strauss 2008, pA01).

Guiso et al (2008) have looked at gender gaps in the 2003 Programme for International Student Assessment, PISA (OECD 2008), and have shown that gaps persist in many countries. In mathematics, girls' performance generally lags behind boys' (an exception being Iceland) and in all countries girls outperform boys in reading. But a key finding is that, although boys' under-performance relative to girls in reading persists, girls' under-performance in mathematics virtually disappears in more gender-equal cultures (as measured by several indicators including the World Education Forum's Gender Gap Index, GGI). This cultural effect, though different in nature to the effects related to minority groups, adds to the view that assessment of performance is more an indicator and less a cause of disadvantage and gaps in levels of achievement between groups.

For quite some time, dubious gender disadvantages in using computer-based assessment have been claimed for girls (eg technology-averse) and boys (eg less competent typists). However, it would appear that such differences and disadvantages as may exist may be disappearing. Indeed Horne's small-scale study (2007) suggests that computerised versions of literacy tests may not be prone to gender effects, producing similar results for boys and girls while paper versions favoured girls. Developments in computer-based assessment are very dynamic, with advances and innovation constantly ongoing. Much research is therefore needed to conclude whether gender effects do transfer from conventional assessment modes or are eliminated in the 'neutral' computerised environment.

Children's rights and assessment

The United Nations Convention on the Rights of the Child (UNCRC 1989) sets out a series of articles designed to give protection to children (defined as young people from birth to age 18) in various aspects of their lives. At least three of these are relevant to making assessment more just, namely:

_ Article 3: In all actions concerning children, the best interests of the child shall be a primary consideration.

_ Article 12: The right of the child to express their views freely in all matters affecting them, their views being given due weight in accordance with their age and maturity.

_ Article 29: The education of the child shall be directed to the development of the child's personality, talents and mental and physical abilities to their fullest potential.

There is very little reported work in the area of assessment and children's rights, but Smith (2006) raises a number of arguments relevant to the field. Working in an Israeli context, she argues that teachers "...violate the selected paragraphs [3, 12, 13 and 29] by current assessment practice" but that they do this "...out of ignorance of the Convention and out of the culture of assessment practised in schools" (p20).

Smith quotes research that shows that schools, under the strictures of the No Child Left Behind Act in the US, will act to exclude the weakest students whose test results would have a negative impact on the average test score, concentrating instead on those who can raise the average performance. Hursh (2005) uses Gilborn and Youdell's (2000) educational 'triage' concept to describe this process of concentrating resources on those who might just pass the tests and leaving those who they judge will definitely fail, and those who will do OK, more or less to their own devices. The reason schools might do this is to make sufficient 'Annual Yearly Progress', a high-stakes situation that can determine funding rewards or penalties for the schools and school districts involved. High school completion rates tell a story of how existing disadvantage is compounded by the additional disadvantage caused by standardised testing. One example is provided by New York's educational policy makers who believed that basing improvements on standardised testing was "...the only way to ensure that all students, including those of colour and those living in poverty, have an opportunity to learn" (Hursh 2005, p609).

The policy, instituted in the mid-1990s, included a requirement to pass five standardised tests in order to graduate. Hursh (2005) questions the efficacy of such programmes, quoting a 17% increase in drop-outs between 1998 and 2000. He noted graduation rates of only 35% for African-American and 31% for Hispanic students, and the damning statistic that English language learners had gone from the highest diploma-earning minority in 1996 to the highest drop-out minority in 2002. Disabled students also showed an increase of 27% in drop-outs in the period 1996-2001.

However misguided it might be - by the rhetoric that higher standards will produce higher levels of outcome through greater efforts by the teachers and students alike - such an assessment process cannot be in the child's best interests (Article 3). And rarely would the child's opinions on the issue be sought or given due attention (Article 12). This would also apply to the design and delivery of the tests, as these processes are simply 'done' to students with all of the decisions being taken by the teacher or by authorities beyond the teacher.

In addition, much summative and standardised testing would undoubtedly fall foul of Article 29, which stipulates that education should help all children to develop their personalities and talents to the fullest potential. Smith (2006) argues that there is a failure to acknowledge the possibilities of multiple intelligences in educational assessment and testing systems. This in turn has led to numerous children being unfairly treated through not being offered the opportunity to develop their talents to the fullest potential, simply because their talents do not align with what is tested and testable. The one-size-fits-all approach rules out the assessment of all but the narrowest and most measureable dimensions of a student's skills and accomplishments, especially those in the 'softer' expressive and creative domains.

Performance washback and marking quality

The Assessment Reform Group has argued that there is considerable impact or washback (we choose not to distinguish the two) on students' motivation from repeated bad performances on tests. For example, students cannot avoid the National Curriculum Assessments and repeated poor performances can be demotivating. The group's research (ARG 2002) indicates that the washback can mean that test performance becomes more highly valued than what is being learned; that testing can reduce the self-esteem of lower-achieving students and can make it harder to convince them that they can succeed in other tasks; that constant failure in practice tests demoralises some students and increases the gap between higher and lower achieving students; and that teaching methods may be restricted to what is necessary for passing tests (eg neglect of practical work).

The quality of marking can also be a factor in assessment unfairness, with the possibility of a variety of biases arising when teachers mark their own students' work. There is considerable evidence that harm can be done to students' aspirations by this bias and teachers' lack of assessment skills (ARG 2006). For example, students with neat work might attract higher marks than those with relatively untidy styles, for the same quality of work. While one might propose that this fits the stereotype of teachers' appreciating the neatness from girls more than the untidiness from boys, gender-related assessment issues can be much more complex. Murphy (1991), for example, cited Goddard-Spear's (1983) experience of having the same piece of science writing marked by teachers. When the teachers were told it was girls' work they marked lower than when they were told it was boys' work. Unruly or disaffected students, compared to those perceived to be well behaved and committed, might suffer similarly for the same quality of work. And perhaps most insidious of all is what Bush (2006) described as the "soft bigotry of low expectations", arising largely from teachers' perceptions of students' backgrounds, their capacity to learn and their capacity to succeed.

The group's research (ARG 2002) indicates that the washback can mean that test performance becomes more highly valued than what is being learned

As time goes on, the hope must be that Paul Dressel's mischievous perception of a grade being "...an inadequate report of an inaccurate judgment by a biased and variable judge of the extent to which a student has attained an undefined level of mastery of an unknown proportion of an indefinite material" (Dressel 1983, p12) will become truly a fiction. However, there will always be a need for vigilance against poor standards of assessment (eg see box below).

A primary school headteacher presented two pieces of work on national television in July 2008. They had been returned as part of the National Curriculum Assessment scripts for her pupils. The issue of unfair and highly dubious marking was revealed as follows.

Asked to write a story about a fairground inventor called Pip Davenport, one pupil wrote: "If he wasent doing enthing els heel help his uncle Herry at the funfair during the day. And had stoody at nigh on other thing he did was invent new rides." The child was given eight marks out of 12 for composition.

Another wrote: "Quickly, it became apparent that Pip was a fantastic rider: a complete natural. But it was his love of horses that led to a tragic accident. An accident that would change his life forever. At the age of 7 he was training for a local competition when his horse, Mandy, swerved sideways unexpectedly, throwing Pip on to the ground, paralysed." This pupil was given seven marks out of 12 for composition, while both were awarded five out of eight for sentence structure. (Source: BBC 2008b)

Measures to reduce biased assessment

A combination of the pursuit of fairness and legislated requirements has arguably consolidated guidelines for arranging access to examinations and tests for disabled students, but other guidelines are needed more generally for content and its presentation. Long-standing and highly respected standards (AERA et al 1999) and codes of practice (JCTP 2004) enshrine many expectations for fair treatment and appropriate content. A leading example is ETS' Commitment to Fairness and Equity in Testing (ETS 2008) in which they claim their tests and other products are evaluated to ensure they are not offensive or controversial, do not reinforce stereotypical views of any group, are free of racial, ethnic, gender, socioeconomic and other forms of bias, and are free of content believed to be inappropriate or derogatory towards any group. Ultimately, e-assessment has the same potential as existing assessment systems to ensure that the effects of any potential bias are reduced through protecting a student's personal details from examiner scrutiny (anonymising for race, gender etc), but automated marking will certainly strengthen the perception of equitable treatment.

The fairness of tests may be secured, ostensibly at least, on the basis of checklist processes such as those of ETS, but if tests are not designed to include assessment of a student's breadth of achievement, focusing instead on weaknesses, the effect can be negative and long-lasting. Reay and Wiliam's Hannah (1999) may not be a rare example of how England's National Curriculum Assessments at that time appeared to her to privilege a purely academic interest in spellings and multiplication tables. Prompted by her teacher's view that a good performance on the SATs was vital, the prospect of not doing well led Hannah to believe that she "...will be a nothing" despite being "...an accomplished writer, a gifted dancer and artist, and good at problem solving" (p346); an impact that is arguably in contravention of children's rights.

7. ASSESSMENT AND SOCIAL JUSTICE

The whole theory of education is radically unsound. Fortunately, in England at any rate, Education produces no effect whatsoever.
(Wilde 1895)

Another aspect that we have emphasised is that disadvantage and social injustice do not arise from a person's own attributes

Would that it were true! We have demonstrated that education and educational assessment can be part of the oppressive machine that helps to perpetuate disadvantage for some of our most vulnerable citizens. That said, we have also emphasised that unfair assessment is a relatively small dimension of the social injustice that can be experienced in our society. Much more damning circumstances arise from the impacts of family and community circumstances, and poor educational provision. Clearly people with fundamental physical and mental difficulties are made even more vulnerable in such circumstances.

Another aspect that we have emphasised is that disadvantage and social injustice do not arise from a person's own attributes. They arise from other people's actions, assumptions, ignorance, misconceptions and prejudice. Whether it is their gender, ethnicity or culture, facility with language, lifestyle, religion or their physical and mental well-being, people can be treated unfairly, sometimes in a patronising "We think this is in your best interests" way and sometimes in a mean and unhelpful way. Even in its relatively minor role, can educational assessments mitigate the negative impacts of some human assessors?

Most definitely not – society itself has much to do in ensuring respect and equitable treatment from all assessors. But the assessment process itself must not add to disadvantage and should lessen it wherever possible. As we discussed, it has been standard practice for many years in mainstream school-sector assessments to ensure access to assessment opportunities, with appropriate arrangements for those with physical and mental difficulties. E-assessment, often in combination with assistive technologies such as screen readers, screen magnifiers and haptic input devices, is already functioning well in the area. Access can be facilitated anytime, anywhere, subject to the necessary resources being available. Clearly in circumstances of deprivation, the necessary computer-based and assistive resources may not be available. Perhaps ensuring access to assessment is the easy bit. There are harder questions. For example:

Do we ensure that assessment is a major process in supporting and promoting the learning of disadvantaged students?
No. In the main, national testing (whether compulsory, eg National Curriculum Assessments, or voluntary, eg Highers and GCSEs) remains a largely summative event in students' lives, with little or no feedback to support next steps and improved learning. More or less oblivious to social deprivation and other forms of disadvantage, such modes of assessment can currently do little to address the impact of social injustice. However, misuse of the aggregate results by the media and politicians can do considerable damage by consolidating unfair and inaccurate stereotypes.

Do we ensure that assessment is an enjoyable and motivating experience, improving disadvantaged students' self-esteem and driving up their standards of achievement?
Yes and no. Pedagogically-linked classroom approaches such as assessment for learning, peer and self-assessment, dynamic assessment and dialogic teaching, are capable of raising all students' participation in assessment activities, engaging them deeply in their own learning and ultimately motivating them to higher standards of achievement. These approaches have considerable potential to reach the individual, making their learning important in ways that class or group teaching cannot match. Students with low self-esteem and confidence, both academic and personal, and those with a lack of motivation or commitment, can benefit greatly from the personalisation offered by some of these techniques. Some modes of summative assessment can also foster high levels of student engagement, for example project and portfolio work.

E-assessment offers considerable enhancements in the capability of assessment agencies to offer feedback electronically; immediately through automated marking systems and relatively promptly in human-assessed systems

In the main, however, the answer is currently no for summative assessment.

Summative assessment at present equates to one-off assessments - with positive impact if the results are good and potentially demotivating impacts if they are not. It is unlikely at present that the reasons for a poor performance would be revealed sufficiently to guide improvements. Increasing moves to return annotated scripts will go some way to addressing this by providing students with the opportunity to evaluate their performance and identify for themselves how improvements might be made.

E-assessment offers considerable enhancements in the capability of assessment agencies to offer feedback electronically; immediately through automated marking systems and relatively promptly in human-assessed systems. Feedback in itself, of course, has the potential to motivate or demotivate students, and the relative ease with which e-assessment can deliver it is therefore something of a double-edged sword. Couched appropriately in constructive language, even a negative message can become formative and helpful, but such an aspiration is as difficult to achieve in a computer-based vehicle as in any conventional oral or paper-based context. It very much depends on who supplies the narrative.

Do we ensure that the content of examinations is appropriate; for example: non-oppressive, meaningful and accessible to all, and non-offensive?
For the most part yes, but inappropriate issues do creep through. These are largely unintended and arise from ignorance and lack of sensitivity to cultural or other features of some members of the test-taking group. An example of this is cited by Newton (2005) in which a Key Stage 2 writing task required the students to write a radio advertisement for a new toy. The Royal National Institute for the Deaf did not think this was fair for deaf students. On the whole, assessment agencies police themselves well on these issues, supported by the various codes of practice mentioned above. As Newton (2005) went on to observe, there can be no disagreement on the point of principle but:

> ...no question-writer could aspire to produce a paper which was of precisely equivalent accessibility for all individual students, or all subgroups of students. On one occasion, deaf pupils may find it slightly harder to gain marks; on another, pupils from upper-class homes; on another, pupils whose teachers had failed to cover a certain aspect of the syllabus; and so on indefinitely.
> (p434)

There is still value, however, in exploring how cultural and language sensitivities can be better addressed. One example is the IBRLA approach used in New Zealand (Bishop et al 2001). The focus of the work was the need to ensure that diagnostic assessment tools, used to assess Māori children's achievements in reading, writing and mathematics in the first four to five years of education were aligned with Māori cultural aspirations, preferences and practices. The IBRLA approach is used to raise critical questions about the power and control issues of Initiation (who initiated the development of the diagnostic tool and for what purpose?), Benefits (who specifically was expected to benefit from the use of this tool and in what ways?), Representation (are Māori cultural aspirations, preferences and practices evident in the development, presentation and use of the tools?), Legitimation (what authority does the diagnostic tool have in terms of Māori cultural aspirations?) and Accountability (to whom are the developers and users of the tool accountable?). In the UK, such a process might be prohibitively costly for national assessments (and impractical as such tests obviously change year on year), but standardised diagnostic tests for various sub-groups could benefit from a cultural validation of the IBRLA type.

In e-assessment the same guidance codes apply as for print-based testing but with the addition of industry standards and guidelines to add to the checking of computer-based approaches. These include the BS ISO/IEC 23988 standard (BS 2007) for information technology used in assessments (which was upgraded to international

standard from the original British Standard BS7988:2002). This standard, in common with the other codes of practice, places considerable emphasis on fairness in its advice to e-assessment developers.

Clausen-May (2007) takes a different tack on assessment design and content issues in her critique of the design and accessibility of items in tests for the Programme for International Student Assessment, PISA (OECD 2008) and the Trends in International Mathematics and Science Study, TIMSS (IEA 2008). She argues that the items are often burdened with irrelevant contextualisation, for example, from PISA 2006: "Mei-Ling from Singapore was preparing to go to South Africa as an exchange student. She needed to change some Singapore dollars (SGD) into South African rand (ZAR)". This is 'dense' with irrelevant information compared to her suggested modification: "Mei-ling is from Singapore. She went to South Africa. She changed some Singapore dollars (SGD) into South African rand (ZAR)" (p157). The extra detail in the existing item runs the risk of being misunderstood (cf the two contexts for 'exchange') or indeed not understood by students with special needs. (We might add 'any students' as one person's familiar discourse may be completely unfamiliar to another in today's multi-diverse society.)

Meeting the PISA requirements for translation of items into a variety of national languages is certain to be more challenging as the sophistication of an item's contextualisation increases, making interpretation of the results problematic. Clausen-May's solution is to promote 'Universal Design' for test items (cf Universal Design for Learning, Hitchcock et al 2002) in which irrelevant information is stripped out of the item texts, enabling ease of reading and understanding, and simpler challenges for multi-national translation. Most importantly, such a design modification to the mainstream tests makes the content much more accessible to the students for whom it is designed, those with various special needs. However, the benefits should be shared by all:

> So long as it is not actually unhelpful to others, and it does not change the construct being assessed, the amendment should be made, not just for the pupils for whom it is designed to be helpful, but to the mainstream version of the test that will be taken by the great majority of pupils.
> (Clausen-May 2007, p156)

Clearly such eminently sensible principles can be applied to e-assessment.

What can e-assessment do to promote social justice that traditional assessment cannot?
In terms of the anti-discriminative and anti-oppressive design and content of assessments, and of the biases that might appear in marking processes, e-assessment can and must meet the same quality standards as traditional assessment. However, in comparison to traditional modes of assessment there are additional benefits to be had from e-assessment for students who do not have adequate learning opportunities, are isolated from assessment opportunities through distance or ill-health, have physical access problems, are challenged by cognitive disabilities or have limited proficiency in English. The benefits may be summarised thus:

_ Computer-based approaches can make the administration of traditional examinations and tests more accessible, providing that the delivery technology, and any necessary assistive technology, is available to the test-taker. Whatever the social or physical barriers that may beset students, e-assessment opportunities will be more accessible, online and on-demand.

_ Computer-based approaches can provide a greater degree of personalisation in access to assessment opportunities (anywhere, anytime), to resources (item banks, practice tests) and to prompt feedback designed to support improved performance.

_ Handheld technologies (mobile phones, PDAs etc) and haptic and other input devices will

greatly improve students' opportunities not only to learn but to participate in assessment.

_ Social networking environments and Web 2.0 tools (wikis, blogs etc) will provide a wide range of disadvantaged learners (eg visually, hearing, cognitively and physically impaired self-help groups, cultural and language communities) with greater scope for tuition and support in learning communities. They will also provide a wide range of peer assessment opportunities to foster progress in learning.

_ Parallel or on-demand translation facilities will assist English as an additional language learners to manage assessment tasks and tests on a much more independent basis.

_ Computerised adaptive testing can tailor examination items to the learner's range of achievement, reducing the potential demotivation that repeated failure can cause by ensuring appropriate levels of difficulty in the assessment undertaken.

_ Automated marking systems can replace human assessors in simple test formats, such as multiple-choice, and can support human assessor judgement in more complex formats including essays, thereby contributing to a reduction in potential assessor bias.

_ E-portfolios will continue to develop as vehicles for students' self-expression, enabling them to choose the content and presentation of their learning in new ways and in new areas of learning outcomes, which traditional assessments such as examinations and tests cannot readily do.

Some of these benefits exist now, in their early stages in some cases and well advanced in others. Arguably the sum of these parts could threaten assessment as we know it; who knows what 20 years into the future might hold? Threading throughout our review, but subtly for the most part, may be part of the answer; a notion that some may see as radical in today's terms. It is the increasing trend away from a conception of assessment as a means simply to differentiate people on the basis of measuring what they cannot do or have not achieved in relation to others – Level 5 vs Level 7, C vs A* etc.

The growing trend is to recognise that in ordinary, low-stakes assessment circumstances, assessment should be conceived as an integrated part of learning, providing evidence of what people have achieved and pointing to next steps. There are a number of key learning and assessment-related challenges and tensions to be faced in the near future, and digital developments are likely to make considerable contributions to their resolution. These challenges and tensions arise from the new trends in learning and skills that are attracting considerable attention and which need to be assessed. They may be projected as follows.

Tailored learning – tailored assessment

There will be a growing need to deal with the tension between increasingly student-centred learning provision and core skills across curricula. For example, as a student takes an individualised programme in dance (student-centred), how can they practise and excel in their aspiration to be a member of a dance company (team-working skills)? The main challenge may be in designing a curriculum that can provide for such a combination of personalised and collective learning, but how can assessment be made flexible enough to be on-demand for the student rather than the curriculum? Clearly the web (and current Web 2.0 tools) has much to offer in both personalising learning and its assessment, and in enabling students to address their assessment needs through sophisticated audio, video and communication facilities in digital learning networks.

Borderless learning – borderless assessment

There will also be a growing tension between providing education without borders and developing locally relevant and indeed understood competences of student graduates. German students taking an engineering course given in Edinburgh may not be learning appropriate local building codes for being an architect in Germany. To a large extent this is the status quo, but there is the prospect of web-based, open courseware repositories developing sufficiently to allow institutions to select modules that can directly address local requirements in a global context. Will students in the future be able to design their specific needs into curricula and their attendant assessments? The answer must be yes, and increasingly so as the educational potential of the web continues to develop.

Today's 15 year-old students mostly have very limited curricula; perhaps it is better to say that they have traditional curricula that are chosen for them

Web-based learning environments of the future will be available at any time or place; indeed there may be considerable changes in how the current, normal functions of schools (teaching and learning) are delivered. If students are off-campus in distant locations (for some nations these students could be in different time zones) then online assessments will need to be robustly secured. Distance is generally not a major issue in the UK (though some island communities can suffer serious restrictions on access to conventional schooling), but if the role of schools was to change considerably, what would be the implications of all assessments having to be online? Experience to date with on-demand, continuous assessments suggest that there may be considerable technical and security problems to be surmounted.

No limits on learning – no limits on assessment

In the future there is likely to be an increasing tension between what a student can learn and what their school may be capable of assessing. Today's 15 year-old students mostly have very limited curricula; perhaps it is better to say that they have traditional curricula that are chosen for them. Outside of what is ordained within the boundary of the school curriculum, choice of subject study is currently rare and tightly controlled. However, the trend towards 'new' core skills means that the vehicle for developing them may well diverge considerably from the current staple 'subject' provision. For example, it is not inconceivable that students of the future may develop skills of independent enquiry, self-management, reflective learning, effective participation and creative thinking (QCA 2008a) by choosing to study Japanese language and culture, the eco-system of the lower Amazon and house construction – or any of the huge list of possibilities that the web offers. In addition, they will have access to local and international experts outside the school, vast library resources across the world and online social networking, special interest groups to support their learning. How will schools as we know them face such challenges? We could be cynical and propose that this unlimited learning jamboree will not be permitted. However, should the rhetoric of new learning be matched by action, what then will be the challenges for assessment? Will schools end up restricting students' progress because they do not have the knowledge of content, or expertise in appropriate assessment methods, to assess this new learning? How might such a global menu for learning be met by appropriate means for assessing it?

8. FINAL WORDS

Future assessment systems will be increasingly effective in using digital technology to mitigate the worst effects of disadvantage arising from poverty (eg by addressing access issues through e-learning and e-assessment), language (eg through parallel translation) and various cultural, gender and religious prejudices (eg through automated marking). However, they must also continue to support the 'basics' of literacy and numeracy, hopefully within a positive and motivating design. And they must continue to meet the need for differentiation between test-takers in various high-stakes contexts. In low-stakes situations (arguably the majority) there will be a continuing push to enable the assessment process to promote, motivate and inform learning by enabling the students to show what they can do.

E-assessment is eminently capable of responding to student diversity and individualised needs. It will be a vital element in making assessment in general change its current emphasis from being curriculum-defined and constrained to addressing a much broader range of students' learning. For many students this will be liberating, reducing the anxiety of test-taking and failure, and preventing the waste of talent and skills that do not fit what is tested and testable in current assessment arrangements. The only regret is that such a trend will only be a small contribution to reducing barriers to a positive education system for all, a system that has to contend with almost overwhelming social and physical disadvantage for many students.

9. REFERENCES

(NB All URLs quoted were accessible on 15 September 2008)

AERA, APA and NCME (1999, revised from 1985 version). Standards for Educational and Psychological Testing. American Educational Research Association, American Psychological Association and National Council on Measurement in Education, Washington DC: American Psychological Association

Altshuler, SJ and Schmautz, T (2006). No Hispanic student left behind: the consequences of 'high stakes' testing. Children and Schools 28 (1) 5-14

ARG (2002). Testing, Motivation and Learning. Assessment Reform Group – Publications. www.assessment-reform-group.org

ARG (2006). The Role of Teachers in the Assessment of Learning. Assessment Reform Group – Publications.
www.assessment-reform-group.org

Attali, Y and Powers, D (2008). A Developmental Writing Scale. RR-08-19 Princeton, NJ: Education and Testing Services.
www.ets.org/Media/Research/pdf/RR-08-19.pdf

Atweh, B (2007). International interactions in mathematical education. In U Gellert and E Jablonka (eds), Mathematisation and Demathematisation: Social, Philosophical and Educational Ramifications, 171-186. Rotterdam/Taipei: Sense Publishers

BBC (2008a). Schools still wait for Sats marks. 22 July 2008.
news.bbc.co.uk/1/hi/education/7519108.stm

BBC (2008b). More questions about Sats results. 17 July 2008.
news.bbc.co.uk/go/pr/fr/-/1/hi/education/7511922.stm

Becta (2006). E-portfolios – Process Model. Coventry: Becta.
partners.becta.org.uk/index.php?section=pv&&catcode=_pv_ep_02&rid=13627

Becta (2008). Computers for Pupils. Coventry: Becta.
localauthorities.becta.org.uk/index.php?section=fd&catcode=la_fu_03

Bishop, R, Berryman, M, Glynn, T and Richardson, C (2001). Diagnostic Assessment Tools in Māori Medium Education: A Stocktake and Preliminary Evaluation. New Zealand Ministry of Education.
www.educationcounts.govt.nz/publications/maori_education/20206

Black, P, Harrison, C, Lee C, Marshall B and Wiliam D (2003). Assessment for Learning: Putting it into Practice. Buckingham: Open University Press

Blair, T (1998). The Third Way. Fabian Society: London

Boston, K (2008). Exam chief calls for on-screen marking. 15 July 2008.
www.kablenet.com/kd.nsf/Frontpage/490291F8697317758025748700 3D3506?OpenDocument

Boyle, A (2005). Sophisticated tasks in e-assessment: What are they? What are their benefits? Proceedings of the 9th Computer Assisted Assessment (CAA) Conference 2005. www.caaconference.com/pastConferences/2005/proceedings/

BS (2007). BS ISO/IEC 23988:2007 A Code of Practice for the Use of Information Technology (IT) in the Delivery of Assessments. www.standardsuk.com/shop/products_view.php?prod=40936

Bush, WG (2006). White House Press Release. 20 July 2006. www.whitehouse.gov/news/releases/2006/07/20060720.html

Camara, WJ and Lane, S (2006). A historical perspective and current views on the standards for educational and psychological testing. Educational Measurement, Issues and Practice 25 (3) 35-41

Carter, CR (1997). Assessment: shifting the responsibility. Journal of Secondary Gifted Education 9 (2): 68-75

Clausen-May, T. International mathematics tests and pupils with special needs. British Journal of Special Needs, 34 (3) 154-161

Cribb, A and Gewirtz, S (2003). Towards a sociology of just practices; an analysis of plural conceptions of justice. In C Vincent (ed), Social Justice Education and Identity 15-29. London: Routledge Falmer

Cumming, JJ and Dickson, EA (2007). Equity in assessment: discrimination and disability issues from an Australian legal perspective. Education and the Law 19 (3/4) 201-220

DCSF (2008). The Standards Site – FAQs. London: Department for Children, Schools and Families. www.standards.dfes.gov.uk/sie/si/SCC/faqs

DDA (1995). Disability Discrimination Act 1995. Office of Public Sector Information (OPSI). www.opsi.gov.uk/acts/acts1995/ukpga_19950050_en_1

DDA (2005). Disability Discrimination Act 2005. Office of Public Sector Information (OPSI). www.opsi.gov.uk/acts/acts2005/ukpga_20050013_en_1

Demie, F (2005). Achievement of Black Caribbean pupils: good practice in Lambeth schools. British Educational Research Journal, 31 (4) 481-508

Demie, F and Strand, S (2006). English language acquisition and educational attainment at the end of secondary school. Educational Studies 32 (2) 215-231

DFES (2007). Gender and Education: the Evidence from England. Department for Education and Skills London: DFES. www.standards.dfes.gov.uk/genderandachievement/pdf/7038_DfES_Gender_Ed.pdf

Dressel, P (1983). Grades: one more tilt at the windmill. In AW Chickering (ed), Bulletin. Memphis: Memphis State University

Duffield, J, Allan, J, Turner, E and Morris, B (2000). Pupils' voices on achievement: an alternative to the standards agenda. Cambridge Journal of Education, 30 (2) 263-274

Elwood, J (2005). Gender and achievement: what have exams got to do with it? Oxford Review of Education, 31 (3), 373-393

ETS (2006). Overview of the Revised GRE Test, Education and Testing Service. www.coe.ilstu.edu/ci/graduate/masters/curriculum_instruction/documents/gre_revised_overview.pdf

ETS (2008). Committed to Fairness and Equity in Testing. Educational Testing Service, www.ets.org/portal/site/ets/menuitem.22f30af61d34e9c39a77b13bc3921509/?vgnextoid=28d65784623f4010VgnVCM10000022f95190RCRD

FairTest (2008). College Board reports confirm: 'new' SAT is not significantly better predictor than old exam, test revision ignored gender, race, language bias concerns; more schools likely to drop admissions testing requirements. Press release of the National Center for Fair and Open Testing, 17 June 2008. www.fairtest.org/fairtest-reacts-new-sat-validity-studies

Fitch, J (1898). Lectures on Teaching. (lectures first given in 1880). Cambridge: Cambridge University Press

Fraser, N (1997). Justice Interruptus: Critical Reflections on the 'Postsocialist' Condition. New York: Routledge

Fusarelli, LD (2004). The potential impact of the No Child Left Behind Act on equity and diversity in American education. Education Policy 18 (1) 71-94

GL (2008). Outlook, January 2008. Granada Learning, London: Chiswick

Gordon, EW (1995). Toward an equitable system of educational assessment. Journal of Negro Education 64 (3) 360-372

Gray, J, Peng, W-J, Steward, S and Thomas, S (2004). Towards a typology of gender-related school effects: some new perspectives on a familiar problem. Oxford Review of Education, 30 (4) 529-550

Guiso, L, Monte, F, Sapienza, P and Zingales, L (2008). Culture, gender and math. Science 320 1164-1165

Harlen, W (2006). On the relationship between assessment for formative and summative purposes. In Gardner, J (ed), Assessment and Learning. 103-117 London: Sage

Hayward, EL (2007). Curriculum, pedagogies and assessment in Scotland: the quest for social justice. 'Ah kent yir faither'. Assessment in Education 14 (2) 251-268

Horne, J (2007). Gender differences in computerised and conventional educational tests. Journal of Computer Assisted Learning, 23 47-55

Hursh, D (2005). The growth of high stakes testing in the USA: accountability, markets and the decline in educational equality. British Educational Research Journal, 31 (5) 605-622

IEA (2008). Trends in International Mathematics and Science Study. International Association for the Evaluation of Educational Achievement.
www.iea.nl

IHE (2007). ETS aborts new GRE. Inside Higher Education.
www.insidehighered.com/news/2007/04/03/ets

JCTP (2004). Code of Fair Testing Practices in Education. Joint Committee on Testing Practices, Washington, DC.
www.apa.org/science/FinalCode.pdf

JISC/QCA (2007). Effective Practice with e-Assessment. Joint Information Systems Committee and Qualifications and Curriculum Authority.
www.jisc.ac.uk/media/documents/themes/elearning/effpraceassess.pdf

Johnston, J and McClune, W (2000). Pupil motivation and attitudes - self-esteem, locus of control, learning disposition and the impact of selection on teaching and learning. In The Effects of the Selective System of Secondary Education in Northern Ireland: Research Papers Volume II. Bangor, Co Down: Department of Education, pp1-37

Kelly, DM and Brandes, GM (2008). Equitable classroom assessment: promoting self-development and self-determination. Interchange 39 (1) 49-76

Keyes, D (2007). Classroom caste system. Washington Post, 9 April, A13.
www.washingtonpost.com/wp-dyn/content/article/2007/04/08/AR2007040800925.html

Kozol, J (1991). Savage Inequalities: Children in America's Schools, New York: Crown Publishers Inc

Leitch, J, Gardner, J, Mitchell, S, Lundy, L, Odena, O, Galanouli, D and Clough, P (2007). Consulting Pupils on the Assessment of their Learning (CPAL) Research Report.
www.cpal.qub.ac.uk

Leonard, M and Davey, C (2001). Thoughts on the 11 plus. Belfast: Save the Children Fund

Levacic, R and Marsh, A (2007). Secondary modern schools: are their pupils disadvantaged? British Educational Research Journal, 33 (2) 155-178

Levacic, R and Woods, PA (2002). Raising school performance in the league tables (part 1): disentangling the effects of social disadvantage. British Educational Research Journal, 28 (2) 207-226

Lynch, K and Baker, J (2005). Equality in education: an equality of condition perspective. Theory and Research in Education, 3 (2) 131-164

Maguire, M, Wooldridge, T and Pratt-Adams, S (2006). The Urban Primary School. London: Open University Press

Mansell, W (2007). Education by Numbers: The Tyranny of Testing. London: Politico's

Mason, R, Pegler, C and Weller, M (2004). E-portfolios: an assessment tool for online courses. British Journal of educational Technology 35 (6) 717-727

McCullough, L (2004). Gender, context and physics assessment. Journal of International Women's Studies, 5 (4) 20-30

McGuire, L (2005). Assessment using new technology. Innovations in Education and Teaching International. 42 (3) 265-276
www.quotationspage.com/quote/282.html

Monaghan, W and Bridgeman, B (2005). e-Rater as Quality Control on Human Scores. Educational Testing Service.
www.ets.org/Media/Research/pdf/RD_Connections2.pdf

Murphy, P (1991). Assessment and gender. Cambridge Journal of Education, 21 (2) 203-213

NCLB (2002). No Child Left Behind Act. The White House.
www.whitehouse.gov/news/releases/2002/01/20020108.html

Newton, PE (2005). The public understanding of measurement inaccuracy. British Educational Research Journal, 31 (4) 419-442

Nichols, P (2005). Evidence for the Interpretation and Use of Scores from an Automated Essay Scorer. Research report 05-02, Pearson Educational Measurement.
www.pearsonedmeasurement.com/downloads/research/RR_05_02.pdf

Nichols, SL, Glass, GV and Berliner, DC (2005). High Stakes Testing and Student Achievement: Problems for the No Child Left Behind Act. Arizona State University: Educational Policy Studies Laboratory.
epicpolicy.org/files/EPSL-0509-105-EPRU.pdf

OECD (2008). Programme for International Student Assessment. Organisation for Economic Cooperation and Development.
www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html

OUT-LAW (2007). Computer-based exam discriminated against blind candidate. Out-Law News, 24 January 2007.
www.out-law.com/default.aspx?page=7692

Pitoniak, MJ and Royer, JM (2001). Testing accommodations for examinees with disabilities: a review of psychometric, legal and social policy issues. Review of Educational Research, 71 (1) 53-104

Plummer, DL (2003) (ed). Handbook of Diversity Management. Lanham, MD: University Press of America

Pressey, SL (1926). A simple apparatus which gives tests and scores - and teaches. School and Society, 23 (586), 373-376

Putwain, D (2007). Test anxiety in UK schoolchildren: prevalence and demographic patterns. British Journal of Educational Psychology, 77 579-593

QCA (2004). A Proposed Blueprint for Delivering e-Assessment. Qualifications and Curriculum Authority.
www.qca.org.uk/libraryAssets/media/6995_blueprint_for_e-assessment.rtf

QCA (2007). E-Assessment: Guide to Effective Practice. Qualifications and Curriculum Authority (England), Council for Curriculum, Examinations and Assessment (Northern Ireland), Scottish Qualifications Authority (Scotland) and the Department for Children, Education, Lifelong Learning and Skills (Wales).
www.qca.org.uk/libraryAssets/media/qca-07-3209_e-Assessment_Guide.pdf

QCA (2008a) The Diploma: an Overview of the Qualification. London: Qualifications and Curriculum Authority.
www.qca.org.uk/libraryAssets/media/The_Diploma_An_overview_of_the_qualification_v3_March_2008.pdf

QCA (2008b). Assessing Pupils' Eligibility for Access Arrangements. London: Qualifications and Curriculum Authority.
www.qca.org.uk/libraryAssets/media/assessing_pupils_eligibility.pdf

QCA (2008c). KS2 Assessment and Reporting Arrangements 2008, National Curriculum Assessments. London: Qualifications and Curriculum Authority.
www.qca.org.uk/eara/documents/KS2_v07aw-2.pdf

Reay, D (2006). The zombie stalking English schools: social class and educational inequality. British Journal of Educational Studies, 54 (3) 288-307

Reay, D and Wiliam, D (1999). 'I'll be a nothing': structure, agency and the construction of identity through assessment. British Educational Research Journal, 25 (3) 343-354

Rawls, J (1971). A Theory of Social Justice. Harvard: Belknap Press

Remedios, R, Ritchie, K and Lieberman, DA (2005). I used to like it but now I don't: the effect of the transfer test in Northern Ireland on pupils' intrinsic motivation. British Journal of Educational Psychology, 75 435-452

Ripley, M (2007). E-assessment – an Update on Research, Policy and Practice. Report 10 Update, Bristol: Futurelab.
www.futurelab.org.uk/resources

SENDA (2001). Special Educational Needs and Disability Act 2001. Office of Public Sector Information (OPSI).
www.opsi.gov.uk/ACTS/acts2001/ukpga_20010010_en_1

SENDO (2005). The Special Educational Needs and Disability (Northern Ireland) Order 2005. Office of Public Sector Information (OPSI).
www.opsi.gov.uk/si/si2005/20051117.htm

Slavin, RE and Madden, NA (2006). Reducing the gap: Success for All and the achievement of African American students. Journal of Negro Education 75 (3) 389-400

Smith, K (2006). Assessment in the light of the UN Convention on the rights of the Child. Paper presented at the NERA 34th Congress, Orebro, 2006. www.ibl.liu.se/content/1/c6/07/77/58/Smith20061.pdf

Stocking, ML, Smith, R and Swanson, L (2000). An Investigation of Approaches to Computerising the GRE Subject Tests. Princeton, NJ: Education and Testing Services. tinyurl.com/3be9a9 or www.ets.org/portal/site/ets/menuitem.c988ba0e5dd572bada20bc47c3921509/?vgnextoid=47f0af5e44df4010VgnVCM10000022f95190RCRD&vgnextchannel=e15246f1674f4010VgnVCM10000022f95190RCRD

Strand, S (2008). Minority Ethnic Pupils in the Longitudinal Study of Young People in England. Cedar Newsletter 20 (Spring) 1 University of Warwick: Centre for Educational Development, Appraisal and Research. Full report at: www.dcsf.gov.uk/research/data/uploadfiles/dcsf-rr002.pdf

Strauss, V (2008). No crisis for boys in schools, study says academic success linked to income. Washington Post, May 20, A01. www.washingtonpost.com/wp-dyn/content/article/2008/05/19/AR2008051902798.html

Taysum, A and Gunter, H (2008). Researching social justice and school leadership in England. Education Citizenship and Social Justice, 3 (2), pp183-199

Teachernet (2008). Computers for Pupils and Access to Technology at Home: Guidance for LAs and Schools. www.teachernet.gov.uk/_doc/10284/CfP_Guidance_May08.doc

Techdis (2008). JISC Techdis Technology Database. Joint Information Systems Committee. www.techdis.ac.uk/gettechnologydatabase

Thomas, RR (1996).. Redefining Diversity. New York: Amacom

Thoreau, HD (1854). Walden. UK: Collector's Library (2004), CRW Publishing Ltd 2004

Thrupp, M and Lupton, R (2006). Taking school contexts seriously: the social justice challenge. British Journal of Educational Studies, 54 (3) 308-328

Thrupp, M and Tomlinson, S (2005). Introduction: education policy, social justice and 'complex hope. British Educational Research Journal, 31 (5) 549-557

Tomlinson Report (2004). 14-19 Curriculum and Qualifications Reform: Final Report of the Working Group on 14-19 Reform. London: Department for Education and Skills. www.dcsf.gov.uk/14-19/documents/Final%20Report.pdf

Torrance, H (2003). Assessment of the National Curriculum in England. In Kellaghan, T and Stufflebeam, DL (eds), International Handbook of Educational Evaluation. Dordrecht: Kluwer Academic Publishers, 905-928

Twining, P, Broadie, R, Cook, D, Ford, K, Morris, D, Twiner, A and Underwood, J. Educational Change and ICT: an Exploration of Priorities 2 and 3 of the DFES Strategy in Schools and Colleges – the Current Landscape and Implementation Issues. Coventry: Becta.
partners.becta.org.uk/page_documents/research/educational_change_and_ict.pdf

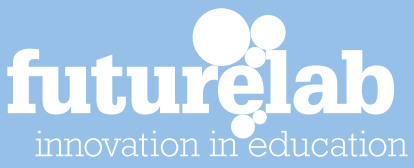
UNCRC (1989). United Nations Convention on the Rights of the Child. New York: United Nations.
www.unhchr.ch/html/menu3/b/k2crc.htm

Walters, S (2007). How do you know that he's bright but lazy? Teachers' assessments of Bangladeshi English as an Additional Language pupils in two Year Three classrooms. Oxford Review of Education, 33 (1) 89-101

Way, WD, Davis, LL and Fitzpatrick, S (2006). Practical Questions in Introducing Computerised Adaptive Testing for K-12 Assessments. Research Report 05-03, Pearson Educational Measurement.
www.pearsonedmeasurement.com/downloads/research/RR_05_03.pdf

Wilde, O (1895). The Importance of Being Earnest. Lady Bracknell, Act 1

Woodfield, R, Earl-Howell, S and Solomon, L (2005). Gender and mode of assessment at university: should we assume female students are better suited to coursework and males to unseen examinations? Assessment and Evaluation in Higher Education, 30 (1) 35-50



Futurelab
1 Canons Road
Harbourside
Bristol BS1 5UH
United Kingdom

tel: +44 (0)117 915 8200
fax: +44 (0)117 915 8201
e-mail: info@futurelab.org.uk
blog: flux.futurelab.org.uk
www.futurelab.org.uk

Registered charity 1113051