

Fairness in multicultural assessment systems

Gordon Stobart*

University of London, UK

This paper is about fairness (equity) in large-scale assessment systems within multicultural societies. It makes the key assumptions that fairness is fundamentally a sociocultural, rather than a technical, issue and that fair assessment cannot be considered in isolation from both the curriculum and the educational opportunities of the students. Equity is defined as a qualitative concern for what is just. This involves, but is not the same as, equality of opportunity and of outcome. In relation to large-scale assessment four topics are addressed: the nature of the assessment system; recognizing experiences of different groups; cultural diversity; and monitoring group performance. The conclusion is that, while we can never achieve fair assessment, we can make it fairer. At the heart of this improvement process is openness about design, constructs and scoring which brings out into the open the values and biases of the test design process.

Fair assessment

No small topic this, so this article is restricted to a general treatment of what, in a multicultural society, 'fair assessment' might mean and its implications for assessment systems such as national tests and examinations.¹ It does not deal with fairness in relation to day-to-day classroom and school assessment, though the issues are much the same. I draw heavily on the contributions to the 26th International Association for Educational Assessment (IAEA) conference in Jerusalem (2000), which had the theme *Educational Assessment in a Multicultural Society*, and on a paper in response to this at the 2003 IAEA conference in Manchester.²

The key assumption underlying this argument is *that fairness is fundamentally a sociocultural, rather than a technical, issue*. Fair assessment cannot be considered in isolation from both the curriculum and the educational opportunities of the students. This may seem obvious, but in developing assessments the quality of schooling and curriculum are often treated as 'givens' about which little can be done. Even if these

*Institute of Education, University of London, 20 Bedford Way, London WC1H 0AL, UK. Email: g.stobart@ioe.ac.uk

are someone else's responsibility, if we have groups that are disadvantaged in this way how do they impact on the fairness of our assessments? Gee (2003) has put this starkly: 'If two children are being assessed on something they have not had equivalent opportunities to learn, the assessment is unjust' (p. 28). This link with curriculum becomes even more critical when assessment is used as an indirect form of curriculum control. For example, Airasian (1988) claims that the growth of high stakes testing in the USA represents a symbolic central control over an education system that does not have an agreed curriculum. This was echoed by Shohamy (2000), who observed:

It is often claimed that hegemonic groups (especially educational bureaucrats) *use* tests for power and control to suppress and eliminate multicultural differences in covert ways. It is the realization by those in authority that test takers will change their behaviors in order to succeed on tests that leads them to introduce tests as disciplinary tools...tests become turned into arms of policy and for imposing control. (p. 3)

Fairness and equity

I am using *fairness* in this paper to capture what is represented by the more technical term *equity*. The legal meaning of this is 'the spirit of justice'. Walter Secada (1989), whose approach to equity I have adopted, elaborates:

Equity attempts to look at the justice of a given state of affairs, a justice that goes beyond acting in agreed upon ways and seeks to look at justice of the arrangements leading up to and resulting from those actions. Note that the actions *do* represent an effort at being just, much as the laws represent our legal and written efforts at justice. However, just as law cannot anticipate all eventualities...so too our actions and social arrangements may be found wanting, in the face of both unforeseen occurrences and our evolving notion of what is just. Equity inhabits this ground between our actions on the one hand and our notions of justice on the other. (p. 81)

Equity, in this definition, is *a qualitative concern for what is just*. While the argument here focuses on assessment practices, this approach also fits with more econometric definitions of equity that focus on the allocation of educational resources and opportunities (e.g. Orfield, 1999).

Equity and equality

The implication is that equity is *not* the same as equality. Equity represents the judgement about whether equality, be it in the form of opportunity and/or of outcomes, achieves just ('fair') results. Equality is essentially a quantitative approach to differences between groups. If we are considering fairness, then the issue of one group performing differently from another cannot be resolved simply at a quantitative level.

For example, it is possible in relation to fairness to have similar outcomes for two groups and yet to see this as unfair to one of the groups, which may have been disadvantaged in terms of access to the curriculum. Similarly it is possible to have unequal group outcomes that may be seen as fair. An example would be where there are group

differences in the application to learning and preparation where each had similar resources and opportunities. The philosopher John Wilson (1991) has argued:

Education is not (only) something that can simply be *given* to people and distributed equally or unequally, like cake. To be educated is not just to have received something but also to have done something...there is always what we may call the question of uptake: whether the individual makes use of whatever opportunities or resources he may be given (p. 223).

Whose fairness is it?

Assessment is a socially embedded activity which can only be understood by taking account of the cultural, economic and political contexts within which it operates (Sutherland, 1996; Gipps, 1999). In looking at the ‘big picture’ sociologists point to the use of assessment as a form of social control so that ‘fair competition’ can actually be a mechanism for mass control in which the system is geared to reward those with ‘cultural capital’ (Bourdieu & Passeron, 1976; Ecclestone & Pryor, 2003). Broadfoot (1996) argues that this form of competition allows those who are not successful to accept their own failure and to acquiesce in the legitimacy of the prevailing social order. She cites intelligence (IQ) testing as a means of social control ‘unsurpassed in teaching the doomed majority that their failure was the result of their own inbuilt inadequacy’ (Broadfoot, 1979, p. 44)—rather than recognizing that IQ testing strongly reflects the different cultural values of the test developers.

The paradox here is that, at the same time that examinations are used for social control and selection and are biased towards those with social capital (typically the middle class), they may be a more equitable means of selection and certification than the systems of patronage they replace. The history of examinations suggests that, while applauding their fairness, little concern was shown for the underlying social bias, for example, that females and certain social groups were not allowed to enter them.

It was this promotion of ‘fair’ selection that was one of the main reasons written examinations were first introduced in England. The naïve nineteenth century enthusiasm for examinations was, according to Sutherland (1992), based on three perceived merits:

- formal examinations were seen as the antithesis of corruption and self-interest. Their claims to neutrality were seen as positive goods, ‘substituting manly self-reliance for cringing place-hunting’ (Dale, 1875, p. 57, quoted in Sutherland (1992)).
- examinations were seen as testing more than attainments or skills—they were perceived as instruments to get at basic abilities;
- ability was equated with merit, talent with virtue.

The written test was seen as having a dual function. ‘Administered to all, it could establish minimum standards, where necessary, raising them. Conducted competitively it offered a means of identifying and rewarding talent’ (Sutherland, 1992, p. 3).

This view has increasingly been challenged, though the assumption that tests can get at basic abilities still seems to prevail. One challenge has come from the USA

where tests were seen to be denying opportunities for advancement, particularly for black students. In the post-1965 Civil Rights legislation era, critics of 'advancement through testing' were pointing out that opportunities to acquire talent, or to be able to show it to sufficient effect in tests and examinations, were not equally distributed (Wood, 1987; Orfield & Kornhaber, 2001). In other words, these tests were biased in favour of the dominant social group.

It is possible for both viewpoints to be advanced by the same individual. I can argue that public tests are important as a means of equalizing opportunities and as a necessary corrective to patronage, while at the same time understanding that tests may be biased in favour of one particular gender, social or ethnic group.

These arguments are largely based on the power relationships between *groups* within a society. Apple (1989) has drawn attention to a political shift in how fairness is defined. His review of public policy in the USA, Britain, and Australia led him to conclude that *equality* had been redefined: it was no longer linked to group oppression and disadvantage, but was concerned to ensure individual choice within a 'free market' perception of the educational community. In Apple's view, this redefinition has reinstated the disadvantage model and underachievement is once again the responsibility of the individual rather than the educational institution, or community:

the current emphasis on 'excellence' (a word with multiple meanings and social uses) has shifted educational discourse so that underachievement is once again increasingly seen as largely the fault of the student. Student failure, which was at least partly interpreted as the fault of severely deficient educational policies and practices, is now being seen as the result of what might be the biological and economic marketplace. (Apple, 1989, pp. 9–10)

Coffield (1999) has echoed this interpretation in relation to the British government's advocacy of 'lifelong learning'. The rhetoric here is that it is the individual's responsibility to continue developing, in his or her own time, the skills that the workplace needs and therefore failure to do this is the individual's, not society's, fault.

Fair assessment, access and curriculum

Fair assessment is inseparable from fairness in access opportunities and in what the curriculum offers. Apple (1989) argues that attention in the equity and education debate must be refocused on important curricular questions, to which Gipps and Murphy (1994) have linked assessment questions (see Table 1). Based on accounts of the difficulties of providing effective education to all in countries such as South Africa (Meier, 2000) and Kenya (Mwachihi & Mbithi, 2000) I have added access and resource questions.

The access questions represent such input variables as students' prior experiences, socio-economic and cultural backgrounds and the educational resources available to them. For example, Meier (2000) reported that in South Africa the teacher-learner ratio was 1:40 for black learners compared to 1:21 for whites. This was compounded by a shortage of qualified teachers in mathematics and science, which meant that many schools for black students did not even offer these subjects, even though they

Table 1. Access, curriculum and assessment questions in relation to equity

Access questions	Curricular questions	Assessment questions
Who gets taught and by whom?	Whose knowledge is taught?	What knowledge is assessed and equated with achievement?
Are there differences in the resources available for different groups?	Why is it taught in a particular way to this particular group?	Are the form, content and mode of assessment appropriate for different groups and individuals?
What is incorporated from the cultures of those attending?	How do we enable the histories and cultures of people of color, and of women, to be taught in responsible and responsive ways? <i>Apple, 1989</i>	Is this range of cultural knowledge reflected in definitions of achievement? How does cultural knowledge mediate individuals' responses to assessment in ways which alter the construct being assessed? <i>Gipps and Murphy, 1994</i>

were part of the official curriculum. Mwachihi and Mbithi (2000) report how in Kenya the introduction of 'cost sharing' has meant that schools now have to fund the purchase of books and other materials, leaving schools in poorer areas without adequate resources. This has been exacerbated by the introduction of a more complex, centrally devised curriculum that is deemed irrelevant to the regional needs. Concerns such as these impact directly on the fairness of the results—something that China has addressed with its rural minority groups by setting differential pass standards on its Higher Education Entrance Examination (Zhao, 2000).

The curricular questions revolve around the cultural capital issues of 'whose curriculum is it?' and are directly linked to the basis and forms of assessment that are discussed next.

Fair assessment in large-scale testing systems

I have chosen four key areas within large-scale testing/examination systems in which to raise issues of fairness, particularly in relation to multicultural societies. These are:

- (1) The nature and requirements of the assessment system itself, for example, how are cultural and linguistic diversity approached?
- (2) How does the content of the assessment reflect the experiences of different groups?
- (3) How do the assessment methods meet the cultural diversity of the candidates?
- (4) How effectively is the performance of different groups monitored and how is this fed back into the system?

Any national assessment system combines political decisions, methods of delivery (often historically based) and the responsibilities of the test agencies. At one extreme are devolved systems in which responsibilities are largely at the school level, there is strong historical continuity and there is minimal government and agency intervention,

for example, Germany (Phillips, 2000). At the other end of the continuum are systems that are politicized, centrally driven and delivered, or strongly regulated, by government agencies, for example, China (Min & Xiuwen, 2001). The balance of these factors will vary, so that the *locus* of decision-making on issues that directly affect fair assessment may also vary. For example, is it the state or the awarding body/testing agency that decides whether tests are offered in more than one language? This section looks at two elements of the system: policy approaches to diversity, including linguistic diversity and issues around equal access.

Assessment systems and approaches to diversity

Shohamy (2000) has proposed three models of how the contributions of different groups are treated:

- (1) *The assimilative model.* In this there is no appreciation of an immigrant's previous knowledge; the task is to master the new knowledge associated with the dominant group. There may be recognition that this takes time to acquire and allowances may be made to ease the process ('pain killers').
- (2) *The recognition model.* In this there is recognition and appreciation of the different knowledge and viewing it as valuable—a situation in which groups are credited for this knowledge and encouraged to maintain it.
- (3) *The interactive model.* In this the knowledge of the 'different' groups affects and influences the dominant group and thus enriches existing knowledge.

While we would probably all aspire to the interactive model Shohamy (2000), representing the highly diverse Israeli culture, is not optimistic:

Even in societies that recognize multiculturalism as part of society there is rarely recognition of the specific and unique knowledge of the different groups in schools. Thus multiculturalism becomes lip-service as there is no 'de facto' recognition of it as legitimate knowledge; educational leaders continue to strive for homogeneous knowledge to be owned by all.

This is even more apparent in educational assessment. In a number of situations there is a gap between curricula and assessment as curricula may, at times, contain statements and intentions for the recognition of diverse knowledge, yet the tests are based on homogeneous knowledge. (p. 3)

Perhaps one litmus test of where an assessment system is in relation to these models is in the attitude to language: how much linguistic diversity does the assessment system reflect? For example, should we:

- assess in only the main language of the culture (for example, USA, England);
- offer the same tests/qualifications in two or more languages (for example, Israel, Wales, Canada; International Baccalaureate Organization);
- use an international language such as French or English as the medium of assessment—rather than the first languages of the culture (for example, West African Examination Council—see Boye & Adeyegbe, 2000)?

Each of these options brings its own benefits and costs. In the monolingual approaches one issue is the accessibility of tests for those who are not using their first language, particularly if this is combined with cultural assumptions in their content (see below). Pollitt *et al.* (2000) provide a case study example of how the monolingual assumptions of mathematics test writers interfered with understanding of an Urdu speaking student taking a mathematics test in English. In Urdu the number of hours 'in a day' (*din*) is 12 (with day-night, *dinraath*, being 24 hours) and there are two words for 'height' (from the ground; of the object)—with both ambiguities capable of generating 'wrong' answers to everyday 'how long would it take...?' and 'how high is...?' questions.

With multilingual approaches the difficulties of comparable 'translation' have been well rehearsed. These range from the issue of whether constructs can be fairly adapted to another culture to variations in demand when words are translated. Hambleton and Patsula (1998) discuss the 'myth' that constructs are universal and consequently all tests can be translated into other languages and adapted for other cultures. Cook (2000) uses their example of speed of response as a key factor in intelligence tests when 'in some cultures speed of response is of minor importance as an operating characteristic for life, and members of these groups often score lower on westernized intelligence tests because of a failure to perform quickly' (p. 3). Cook concludes from her work on the factors affecting the validity on tests given in different languages that 'it is very difficult to adapt a test constructed for a particular population to the language and culture of the second population and maintain construct equivalence between the two tests' (p. 11). Ruddock and Evans (2000) also provide some construct dilemmas in providing English-Welsh translations. These include differences in demand, for example, giving the meaning in science of 'hibernation' may be a less demanding item in Welsh, in which it is translated 'sleeps in winter'.

Equal access

Without fairness in access to resources and in the curriculum it is hard to see how there can be fair assessment. The issue becomes the extent to which, in order not to undermine the validity of the assessment, awarding bodies and test development agencies must concern themselves with the quality of a centre's preparation for an examination or test. One kind of contribution to the 'level playing field' is for awarding bodies to provide centres with sufficient guidance and feedback that they understand the *rules of engagement*. The continuity of the style and form of assessment also means that past papers, examiners' reports and statistical data provide opportunities for centres to prepare themselves effectively. The return of examination scripts to centres could also offer important feedback.

This raises the question of whether this is sufficient, particularly where there may be gross disparities between the levels of competence in centres—especially if the already disadvantaged sections of society may suffer. The West African Examinations Council inspects new centres before approval is given to schools to present candidates

for examinations (Boye & Adeyegbe, 2000). An issue here would be what happens if a centre is found 'unfit' and how then are students given opportunities to access the examination.

This same concern has been raised in the United States with the suggestion that, since opportunity to learn is a key factor in performance, schools may have to '*certify delivery standards*' as part of a system for monitoring instructional experiences (Linn, 1993).

The content of assessment

There is no cultural neutrality in assessment or in the selection of what is to be assessed. This applies as much to mathematics as it does to history, and attempts to portray any assessment as 'acultural' are a mistake. Cumming (2000) observes 'Acultural knowledge has definite cultural roots. This is knowledge that is privileged in our standards and testing procedures' (p. 4). She goes on to raise two key questions which link with those in Table 1:

(1) *When setting standards and test content, are we really sure this is the knowledge we need?* This is essentially a validity issue. I would argue that both our curricula and our examinations are full of things we do not need, in the sense of them having little or no relevance to modern life. They are there for historical reasons and to underpin the belief that particular knowledge is constant and important. Mathematics is an interesting case in point. Cumming points out that while the International Life Skills Survey and the PISA study are looking at 'essential life learnings' we have curricula that see algebra and trigonometry as 'essential learnings' (when did you last use the cosine rule?). Shohamy (2000) provides the example of Ethiopian immigrant students to Israel who are much better than non-Ethiopian students at doing precise approximations and estimations. This useful life skill is not, however, part of the national curriculum so is not likely to be valued.

(2) *Are we really privileging certain knowledges to maintain a dominant culture and in doing so ensuring perpetuation of ourselves, as people who have succeeded in the formal educational culture to date?* An occupational hazard in assessment is the wish to standardize knowledge and minimize differences in understanding, thus encouraging assimilation. History examinations may well, in many cultures, provide a case study of how the content of a curriculum or syllabus is contentious—and how this will be compounded by what is assessed. Whose history is to be studied, and how is it to be interpreted? A powerful example of this dilemma was during the period of Gorbachev's *glasnost* in the Soviet Union. The new openness that this brought meant 'official' Russian history had to be radically re-written, to the extent that for several years there was a moratorium on school history examinations as there was 'no history'. A similar process is going on with the curriculum in South Africa (Meier, 2000).

In England the history programme for the national curriculum proved controversial, with the then prime minister, Mrs Thatcher, actively debating (and changing) the details with her ministers and advisers. This debate continues, particularly in relation to how 'British' this history ought to be. At GCSE and GCE³ the multicultural tensions

have been partly addressed by offering a range of options within a syllabus to reflect the diversity of interest. This in turn raises issues of interpretation and comparability.

The issue for test developers is how they ensure that their sampling of the subject offers opportunities for the different groups who will be taking the test:

We need to encourage clearer articulation of the test/examination developers' constructs on which the assessment is based, so that the construct validity may be examined by test takers and users. Test developers need to give a justification for inclusion of context and types of response mode in relation to the evidence we have about how this interacts with group differences and curriculum experience. (Stobart & Gipps, 1998, p. 48)

Assessment methods

We are now well aware that the form of assessment can differentially affect results for different groups. In England there has been far more analysis of this in relation to gender than to ethnicity. We know that during compulsory schooling (up to 16 years) girls are likely to outperform boys on tasks that involve open-ended writing, particularly when this involves personal response. The gap narrows if the responses are fixed choice or short answer (Gipps & Murphy, 1994), though even within multiple-choice tests there are differential response patterns. Carlton (2000) has shown that in such tests females perform better than males matched for ability on questions in which the content is a narrative or is in a humanities field and when the content deals with human relationships. As the context of an item grows larger the relative performance of females also improves. Males outperform females on questions relating to science, technical matters, sports, war or diplomacy. We also know that where examinations have a coursework element the performance of girls is likely to be more consistent, though the effect this has on final grades has often been overstated (Elwood, 1995).

We know less about other aspects of the form of assessment, particularly in relation to ethnicity. For example, *oral* assessment plays little part in the examination system in England outside examining languages. Does the emphasis on written response disadvantage groups who place more emphasis on oral communication in their culture? Rudduck (1999) has raised this in relation to Afro-Caribbean boys in England who, as a group, often perform less well than others on written examinations.

A similar issue which has attracted little research has been the impact of the 'individualism' of test-taking on groups where the emphasis is on cooperative learning, an issue raised by Erickson over fifty years ago (Erickson, 1950). Sutton (2000) has raised this in relation to Tongan students moving to New Zealand and Cree students in Canada. She observes that 'Much has been done in many ex-colonial countries to return the content of schooling to the cultural roots and history of the original society, but the assessment assumptions remain intact' (p. 27).

Those responsible for assessment systems have to look for manageable and valid ways of meeting such diversity. Some systems may be restricted to a single form of assessment, for example, national tests. Here the effort is on the development of tests that are as fair as possible for the groups taking them. The claim here is that fairness in objective tests is not simply a matter of technical manipulations, for example,

Differential Item Functioning techniques, to equalize outcome for different groups (Goldstein, 1993). Elder (1997), in relation to language testing, concludes that ‘the choice of a criterion against which test bias is estimated and the subsequent interpretation of test scores may be governed by educational or institutional imperatives which are themselves subject to bias’ (p. 272). Developing fair tests is therefore a process that involves social and political assumptions about the construct being assessed and the validity of the items. Gipps (1995) claims that the focus on statistical techniques for elimination of ‘biased’ items ‘not only confounds the construct being assessed, but has distracted attention from wider equity issues such as actual equality in access to learning, ‘biased’ curriculum, and inhibiting classroom practices’ (p. 274).

An aim for less restricted systems may be to encourage the use of a range of assessment modes and tasks so that those who are disadvantaged on one assessment have an opportunity to offer alternative evidence of their expertise (Linn, 1992). This is one purpose of coursework in qualifications in England, of ‘performance assessment’ in the USA and of Teacher Assessment in Sweden and Germany (Gipps & Stobart, 2003). These methods are generally only part of the assessment system, and each brings its own limitations in relation to fairness. For example, Baker and O’Neil (1994) report some uncomfortable findings on performance assessment in relation to ethnic minorities:

the minority community’s perception of the self-evident merit of performance assessment deserves additional exploration...The major assertion was that performance-based assessment reform is a creation of the majority community intended to hold back the progress of disadvantaged children. Performance-based assessment is obviously grounded in a different instructional model, one in which the majority of teachers of disadvantaged children may be unprepared.

Although most of the concerns were articulated by African-Americans, there was also the early recognition that much of performance-based assessment required strong language skills by students to explain or document their accomplishments. It is undeniable that in the US, performance assessment is a white, middle class venture, promoted by high achieving people, disproportionately women. Minority communities must not once again become unwilling recipients of innovations which others believe are good for them. (pp. 13–14)

This concern has been echoed by Meier (2000) in relation to the Outcomes-based Education and Curriculum intended for South African schools because ‘the ignorance among assessors and learners of how to use these methods can again have an impact on assessment results in the sense that some learners may benefit or be harmed depending on the teacher’s interpretation of the methods used’ (p. 10). It is clear from just these few examples that fair assessment in multicultural settings is something we strive for rather than fully achieve. Knowing how well we are doing depends on how we monitor assessment and use this feedback.

Monitoring assessment

If different groups are taking the same, or comparable, tests it seems reasonable to evaluate their relative performances. This could involve the distribution of grades or

scores. It could also involve performance at a component or question level. This is done routinely for gender, which is recorded and is therefore easy to analyse. In England ethnicity is not recorded for either national tests or public examinations. Any research has had to collect the data directly, or rely on some local education authorities that routinely collect it. This situation is compounded if we wish to include socio-economic status, which is often a critical factor in making any comparisons between ethnic groups.

Increasingly more detailed data will become available about performance on tests and how this links to students' background and schooling. For example, in England each pupil now has a unique pupil number (UPN), which means data (including socio-economic information) on national test results and examinations throughout pupils' school careers can be used to look at differential progress. The issue then is the use that is made of such findings in relation to groups of pupils who may not be making the progress expected and how these findings impact on what and how we assess.

Conclusion

We have moved well beyond the naïve assumptions of nineteenth century examiners that a written test under standardized conditions was inherently fair. This paper has considered how fairness (equity) represents a complex qualitative judgement about the interactions of inputs, processes and outcomes. For those involved in large-scale assessment this raises issues about how these systems handle cultural diversity—particularly those of languages, centre accreditation, the content of the assessment and assessment methods.

The argument here has been that fairness is essentially a social process and judgement, 'a qualitative concern for what is just', which is informed by, but not the same as, equality issues. The role of assessment systems in this are critical, not only in terms of how cultural and social differences are recognized and handled, but in the impact on how the curriculum is interpreted and taught. The paper also raised the wider issue of equality in access and the problems for fair assessment if access and opportunities for those involved are very different.

We will never achieve fair assessment, but we can make it fairer:

The best defence against inequitable assessment is openness. Openness about design, constructs and scoring, will bring out into the open the values and biases of the test design process, offer an opportunity for debate about cultural and social influences, and open up the relationship between assessor and learner. These developments are possible, but they do require political will. (Gipps, 1999, p. 385)

Notes

1. I use 'tests' throughout to cover both tests and examinations.
2. *Revisiting Fair Assessment in a Multicultural Society*, International Association for Educational Assessment, Manchester, UK, 5–10 October 2003.

3. The General Certificate of Secondary Education is a subject specific examination usually taken at the end of compulsory schooling (age 16) with the Advanced level General Certificate of Education taken at 18.

Notes on contributor

Gordon Stobart is Reader in Education at the University of London, having spent 15 years working on national tests and examinations for examination boards and government agencies. He previously worked as a secondary school teacher and as an Educational Psychologist after which he studied in the USA as a Fulbright Scholar. He is a member of the Assessment Reform Group, and the editor of *Assessment in Education*.

References

- Airasian, P. W. (1988) Measurement driven instruction: a closer look, *Educational Measurement: Issues and Practice*, Winter, 6–11
- Apple, M. W. (1989) How equality has been redefined in the conservative restoration, in: W. G. Secada (Ed.) *Equity in education* (New York, Falmer Press), 7–25.
- Baker, E. & O'Neil, H. F. (1994) Performance assessment and equity: a view from the USA, *Assessment in Education*, 1(1), 11–26.
- Bourdieu, P. & Passeron, J. C. (1976) *Reproduction in education society and culture* (London, Sage).
- Boye, S. A. M. & Adeyegbe, S. O. (2000) The implications of educational assessment in a multicultural context: WAEC as a case study, paper presented at the 26th IAEA Conference, Jerusalem, 14–19 May.
- Broadfoot, P. (1979) *Assessment, schools and society* (London, Methuen).
- Broadfoot, P. (1996) Assessment and learning: power or partnership, in: H. Goldstein & T. Lewis (Eds) *Assessment: problems, developments and statistical issues* (Chichester, John Wiley), 21–40.
- Carlton, S. T. (2000) Contextual factors in group differences in assessment, paper presented at 26th IAEA Conference, Jerusalem, 14–19 May.
- Coffield, F. (1999) Breaking the consensus: lifelong learning as social control, *British Educational Research Journal*, 25(4), 479–499.
- Cook, L. L. (2000) Factors affecting the validity for scores obtained on tests given in different languages to examinees of different cultural backgrounds, paper presented at the 26th IAEA Conference, Jerusalem, 14–19 May.
- Cumming, J. (2000) After DIF, what culture remains?, paper presented at the 26th IAEA Conference, Jerusalem, 14–19 May.
- Dale, W. (1875) *State of the medical profession in Great Britain and Ireland*, quoted in Sutherland, G. (1992) Examinations, formal qualifications and the construction of professional identities: a British case-study 1880–1940. Paper presented at the IREX—Hungarian Academy of Sciences Conference, Budapest, 17–19 August.
- Ecclestone, K. & Pryor, J. (2003) 'Learning careers' or 'assessment careers'? The impact of assessment systems on learning, *British Educational Research Journal*, 29(4), 471–488.
- Elder, C. (1997) What does test bias have to do with fairness? *Language Testing*, 14(3), 261–277.
- Elwood, J. (1995) Undermining gender stereotypes: examination and coursework performance in the UK at 16, *Assessment in Education*, 2(3), 283–303.
- Erikson, E. H. (1950) *Childhood and society* (New York, W. W. Norton).
- Gee, J. P. (2003) Opportunity to learn: a language-based perspective on assessment, *Assessment in Education*, 10(1), 27–46.

- Gipps, C. (1995) What do we mean by equity in relation to assessment? *Assessment in Education*, 2(3), 271–281.
- Gipps, C. (1999) Sociocultural aspects of assessment, *Review of Research in Education*, 24, 357–392.
- Gipps, C. & Murphy, P. (1994) *A fair test? Assessment, achievement and equity* (Buckingham, Open University Press).
- Gipps, C. & Stobart, G. (2003) Alternative assessment, in: T. Kellaghan & D. L. Stufflebeam (Eds) *International handbook of educational evaluation* (Dordrecht, Kluwer), 549–575.
- Goldstein, H. (1993) Assessing group differences, *Oxford Review of Education*, 19, 141–150.
- Hambleton, R. K. & Patsula, L. (1998) Increasing the validity of adapted tests: myths to be avoided and guidelines for improving tests adaptation practices, paper presented at the *American Educational Research Association (AERA) Conference*, San Diego, 13–17 April.
- Linn, M. C. (1992) Gender differences in educational achievement, in: J. Pfliegerer (Ed.) *Sex equity in educational opportunity, achievement and testing* (Proceedings of the 1991 Educational Testing Service Invitational Conference) (Princeton, NJ, Educational Testing Service), 11–50.
- Linn, R. L. (1993) Educational assessment: expanded expectations and challenges, *Educational Evaluation and Policy Analysis*, 15, 1–16.
- Meier, C. (2000) The influence of educational opportunities on assessment results in a multicultural South Africa, Paper presented at the *26th IAEA Conference*, Jerusalem, 14–19 May.
- Min, H. & Xiuwen, Y. (2001) Educational assessment in China: lessons from history and future prospects, *Assessment in Education*, 8(1), 5–10.
- Mwachihhi, J. M. & Mbithi, M. J. (2000) Assessment and equity assurance in the Kenyan multicultural background, paper presented at *26th IAEA Conference*, Jerusalem, 14–19 May.
- Orfield, G. (1999) Policy and equity: a third of a century of educational reforms in the United States, *Prospects*, XXIX(4), 579–594.
- Orfield, G. & Kornhaber, M. L. (2001) *Raising standards of raising barriers?* (New York, Century Foundation).
- Phillips, D. (2000) Learning from elsewhere in education: some perennial problems revisited with reference to British interest in Germany, *Comparative Education*, 36(3), 297–307.
- Pollitt, A., Marriott, C. & Ahmed, A. (2000) Language, contextual and cultural constraints on examination performance, paper presented at the *26th IAEA Conference*, Jerusalem, 14–19 May.
- Ruddock, G. & Evans, S. W. (2000) Working in two languages (or two variants of the same language), paper presented at the *26th IAEA Conference*, Jerusalem, 14–19 May.
- Ruddock, J. (1999) *Compound fractures*, seminar presentation, QCA, London.
- Secada, W. G. (1989) Educational equity versus equality of education: an alternative conception, in: W. G. Secada (Ed.) *Equity in education* (New York, Falmer Press), 68–102.
- Shohamy, E. (2000) Educational assessment in a multicultural society: issues and challenges, paper presented at the *26th IAEA Conference*, Jerusalem, 14–19 May.
- Stobart, G. & Gipps, C. (1998) The underachievement debate: fairness and equity in assessment, *British Journal of Curriculum and Assessment*, 8(3), 43–49.
- Sutherland, G. (1992) Examinations, formal qualifications and the construction of professional identities: a British case-study 1880–1940, paper presented at the *IREX—Hungarian Academy of Sciences Conference*, Budapest, 17–19 August.
- Sutherland, G. (1996) Assessment: some historical issues partnership, in: H. Goldstein & T. Lewis (Eds) *Assessment: problems, developments and statistical issues* (Chichester, John Wiley), 9–20.
- Sutton, R. (2000) Cultural issues in mass student testing, *Management in Education*, 14(1), 27.
- Wilson, J. (1991) Education and equality: some conceptual questions, *Oxford Review of Education*, 17(2), pp. 223–230.
- Wood, R. (1987) *Assessment and equal opportunities*, public lecture at ULIE, 11 November.
- Zhao, H. (2000) The minority nationality related issues in China public examinations, paper presented at the *26th IAEA Conference*, Jerusalem, 14–19 May.