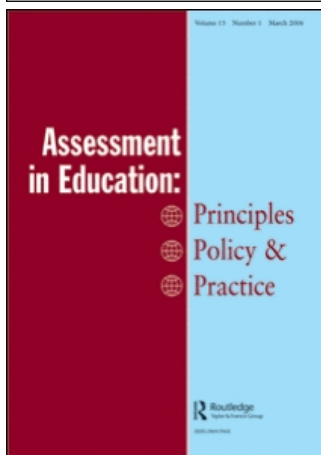


This article was downloaded by:[Swets Content Distribution]  
On: 11 December 2007  
Access Details: [subscription number 768307933]  
Publisher: Routledge  
Informa Ltd Registered in England and Wales Registered Number: 1072954  
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Assessment in Education: Principles, Policy & Practice

Publication details, including instructions for authors and subscription information:  
<http://www.informaworld.com/smpp/title-content=t713404048>

### Redefining assessment? The first ten years of **assessment in education**

Patricia Broadfoot <sup>a</sup>, Paul Black <sup>b</sup>

<sup>a</sup> University of Bristol, UK

<sup>b</sup> King's College London, UK

Online Publication Date: 01 March 2004

To cite this Article: Broadfoot, Patricia and Black, Paul (2004) 'Redefining assessment? The first ten years of **assessment in education**', *Assessment in Education: Principles, Policy & Practice*, 11:1, 7 - 26

To link to this article: DOI: 10.1080/0969594042000208976

URL: <http://dx.doi.org/10.1080/0969594042000208976>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Redefining assessment? The first ten years of *Assessment in Education*

Patricia Broadfoot<sup>1\*</sup> & Paul Black<sup>2</sup>

<sup>1</sup>University of Bristol, UK; <sup>2</sup>King's College London, UK

The completion of the first ten years of this journal is an occasion for review and reflection. The main issues that have been addressed over the ten years are summarized in four main sections: Purposes, International Trends, Quality Concerns and Assessment for Learning. Each of these illustrates the underlying significance of the themes of *principles*, *policy* and *practice*, which the journal highlights in its subtitle. The many contributions to these themes that the journal has published illustrate the diversity and complex interactions of the issues. They also illustrate that, across the world, political and public pressures have had the effect of enhancing the dominance of assessment so that the decade has seen a hardening, rather than any resolution, of its many negative effects on society. A closing section looks ahead, arguing that there is a move to rethink more radically the practices and priorities of assessment if it is to respond to human needs rather than to frustrate them.

## Introduction

In 1993, a new international journal was launched. Its title was *Assessment in Education: principles, policy and practice*. The instigators were a team of academics from the Graduate School of Education, University of Bristol, UK and from the Institute of Education, University of London. This team covered a wide range of disciplines, from the technical to the social. The initiative reflected the lack of a mainstream academic journal devoted to the dissemination of all aspects of research on educational assessment. Whilst there were already in existence a number of long-established international journals dealing with many of the *technical* aspects of testing, there was no journal that focused more broadly on the policy and practice of assessment around the world. Given the unprecedented growth in educational assessment of all kinds in the decade or so leading up to 1993, the lack of a dedicated voice for disseminating the substantial volume of international research in this field was a significant omission. It represented a barrier to the development of greater international understanding and insight concerning the impact of different forms of assessment on educational policy and practice and about the ways in which both might be developed better to meet their intended purposes.

Thus *Assessment in Education* was launched. As the journal's subtitle implies, its aim was to provide a forum for scholarly discussion of issues of *principle*, *policy* and

---

\*Corresponding author: Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol BS8 1JA, UK. Email: P.M.Broadfoot@bris.ac.uk

*practice* as these were expressed in significant and wide-ranging developments in educational assessment. From the outset, *Assessment in Education* has combined a desire to *inform*—by providing up-to-date and rigorous descriptive material about assessment practices in various parts of the world, including discussions of technical issues—with a desire to *critique*, by providing analyses of educational assessment phenomena that are both original and relevant.

Characteristic of the journal is its awareness of assessment *within its social context*. Whilst the explicit emphasis in this respect varies from article to article, underpinning all the analyses is a recognition that decisions about who and what is to be assessed, for what purpose and by what method, reflect a particular social context. By the same token it is recognized that the consequences of these decisions are likely to be different depending on relativities of time and place. At one extreme, these relativities concern international differences of the broadest kind, between developed and developing countries, for example; at the other they may be embedded in the simplest of interactions—between a teacher and a student in a particular classroom. In each case, however, the underlying principle is the same, namely that educational assessment must be understood as a *social* practice, an art as much as a science, a humanistic project with all the challenges this implies and with all the potential scope for both good and ill in the business of education.

The design of *Assessment in Education* reflects this overall purpose and rationale. As well as pursuing an editorial policy that makes these goals explicit, its contributions include not only conventional academic articles but also major research reviews with invited responses to stimulate debate; special issues devoted to an explicitly international consideration of a particular topic; and extended book reviews which allow leading scholars in the field to offer more general ‘state of the art’ discussions about key topics. In addition, the journal regularly includes ‘country profiles’. These are written according to a standard template by an assessment expert in the particular country being covered who is in a position to offer clear, up-to-date insights about both their national assessment arrangements and a well-informed critique of the key challenges being faced in that particular setting. These elements of the journal are designed to support one of its key goals—disseminating information about the wealth of assessment activity and debate in less well-known parts of the world and especially to the Anglophone world, where such experience can be all too easily ignored.

*Assessment in Education* has now been in existence for ten years. This milestone represents a good opportunity to review the journal’s achievements to date. It also represents a good opportunity to review the field of assessment scholarship—especially as it has been represented in the pages of this particular publication—and to sketch in possible developments that are likely to characterize assessment developments over the next ten years or so.

## Our approach

The field of assessment research is extensive. It is therefore necessary to be selective. For this review we have chosen to concentrate on four key topics: the different

purposes of assessment and the tensions between them; international issues in assessment; quality concerns and assessment for learning. These represent some of the most challenging and cutting-edge aspects of assessment research at the present time. The topics we have chosen emphasize the social, rather than the technical. One of the features of *Assessment in Education*, however, has been its emphasis on situated discussions of technical matters within their social contexts.

Since one of the principal aims of this article is to review the contribution of *Assessment in Education* at the end of its first ten years, we have chosen to explore these four topics by drawing mainly on material published in the journal itself. We are well aware that this constitutes only a small part of the wealth of related research literature that is available on these topics, and we have referred to a few papers published elsewhere where these make a unique contribution to our argument.

Our analysis is linked by three central ideas which are embodied in the journal's subtitle—*principles, policy and practice*. These ideas serve further to emphasize the importance of addressing purpose and effect in the study of educational assessment. With regard to *principles*, we wish to examine how far the search for guiding principles in assessment has been pursued and whether indeed it is either possible or necessary to seek so to do.

The importance of policy speaks for itself. Firstly, decisions about assessment procedures—particularly those concerning 'high-stakes' testing of various kinds—are as often based on perceived *political* appeal as they are on a systematic knowledge of the scientific evidence concerning fitness for purpose. Moreover, although it is possible to trace *policy* issues in assessment back to the earliest days of public examinations when, for example, Napoleon recognized the powerful contribution nationally controlled assessment procedures could play in cementing national unity, in recent years the importance of assessment as a policy tool has grown enormously as governments have increasingly come to realize its powerful potential as a mechanism of state control.

Assessment serves as a communicative device between the world of education and that of the wider society. This spectrum of communication ranges from the most informal of exchanges to the extremely formal, spanning everything from school reports to high-stakes public examinations, and from individual job interviews to national monitoring, the common factor being the use of assessment data of one kind or another as a publicly acceptable code for quality. Closely associated with this is the issue of legitimacy. The results of any particular assessment device must be accorded 'trust' by the public if the consequences are to be acceptable. It is also true, however, that assessment procedures that enjoy public legitimacy may not be subject to the scrutiny that they ought to have.

Thus, assessment policy debates and the scale and significance of recent developments, as they pertain to our four topics, will help to shape the analysis that follows.

The journal's third theme is that of *practice*. This term arguably embraces every aspect of assessment in its concern with delivery, for it is the thinking, the habits, the technologies and the politics of a particular age and time that combine to shape the assessment practices that are realized in schools, colleges and universities, in workplaces and in less formal learning environments. Thus in what follows, we seek to

highlight some of the new insights into assessment practice that *Assessment in Education* has provided.

The first part of this article is essentially retrospective. Its aim is to highlight in four sections, each devoted to a particular topic, some of the most defining themes and issues that have characterized the journal's content over its first ten years. Following these four, the final section of the article offers a more discursive set of reflections on what emerges from this analysis. If by looking back we can discern the more significant insights on which to build for the future, so we may aspire to our overall goal of providing an outline blueprint of what are likely to be the key assessment themes in the years ahead.

### Purposes of assessment

Perhaps the most important aspect of this topic is not the issue of what the specific purpose of any particular assessment activity is, but rather the extent to which the issue of purpose is made overt at all. A great deal of the assessment research literature centres on improving the methodologies of well-established assessment techniques, rather than questioning either the validity of the approach itself or its suitability for the purpose in question.

A necessary first step towards achieving the most effective match between the purpose of a particular assessment and the tools chosen to fulfil that purpose is a clear articulation of the range of different possibilities. Within the familiar lexicon of acknowledged assessment purposes, such as feedback, reporting, certification, selection, accountability and national comparison, various contributions to *Assessment in Education* have helped to identify a more subtle subset of purposes, which serve to underline the pervasive power of assessment to define and shape every aspect of educational life. Some examples which fall within this general theme include: the 1999 special issue on science education (Vol. 6, No. 1; see Appendix 1); Mavromatis' (1997) analysis of the way in which primary school teachers in Greece use assessment as a mechanism for controlling class behaviour and attention, and Rowe and Hill's (1996) exploration of the use of item-response models to track the emerging learning profile of a student. In another example, Harlen and James (1997) highlight the use of assessment for encouraging 'deep' rather than 'surface' learning in higher education, and, as discussed in more detail below, a number of other authors such as Klenowski (1995) and the 1998 special issue on assessment and classroom learning (Vol. 5, No. 1) explore the subtleties of individuals' own involvement in assessment—both learners and teachers—as an influence on the development of their capacity and motivation to learn. From these examples chosen from a range of others, it is clear that the familiar tools of assessment can be employed for a great variety of purposes, some of which are potentially of great educational value but are not currently well understood or even identified, suggesting the need for continued research efforts to explore further untapped potential in the practice of assessment.

This increasingly explicit scholarly and professional awareness of the range of potential purposes of assessment has been matched at a more macro level in the

realm of policy. For example, the 1998 special issue edited by Koretz, Broadfoot and Wolf on portfolios and records of achievement (Vol. 5, No. 3), provides a range of international insights concerning the advent of new policy goals for assessment, whether these are raising achievement, changing the focus of curriculum priorities through assessment, assuring standards, introducing performance management systems for teachers, institutional quality assurance and control, defining 'standards' through the publication of league tables, and so on.

What emerges most powerfully from a consideration of these increasingly explicit assessment purposes is the sheer range of assessment activity; the way in which assessment of one kind or another now penetrates social, corporate and political life. What was once largely the preserve of professional educationists and test developers has become translated into a set of more generic perspectives that can be applied in almost any social setting.

Closely associated with the issue of the scale and impact of assessment and the corresponding need for clear principles and policing for this activity, is that of the tensions between different purposes. Eckstein and Noah's (1993) now classic articulation of the dilemmas surrounding public examination systems and the way in which different countries balance these tensions depending on public tradition, the political context and policy priorities was the subject of a review essay in the inaugural issue of *Assessment in Education* (Sadler, 1994). These tensions have re-emerged recently in the more subtle guise of the apparent incompatibility between policies which lead to an increasingly 'test-driven' educational culture in many parts of the world alongside an explicit policy commitment to encouraging 'lifelong learning'. As the goals of education have expanded in recent years in response to changes in the nature of work and a recognition of the crucial importance of 'lifelong' education for providing a competitive edge in increasingly global economic competition, so educational policy attention in developed countries at least, has begun to focus on how to encourage more young people both to stay on longer within the formal education system and to be equipped and motivated to keep coming back to education throughout their lives.

However, the research findings presented in this journal—see, for example, Crooks, *et al.* (1996), Airasian and Gregory (1997) and Firestone (1998), as well as elsewhere—make it clear that this 'empowerment' agenda cannot be achieved alongside the punitive use of high-stakes testing to raise 'standards'. Any possible short-term gains that the more or less extreme instrumentalism of the latter engenders, encourages 'teaching the test' (Morrison & Tang Fun Hei, 2002) as well as anxiety and low self-esteem among the less successful (Harlen & Deakin-Crick, 2003), and is bought at the price of turning many students off formal learning forever. It would seem from these studies, that whilst it may be technically possible to reconcile formative and summative assessment within the same spirit of personal growth and empowerment, there will need to be a substantial shift in the prevailing political and policy priorities of many countries to achieve this.

## International trends

Perhaps the single most striking conclusion from a review of the contents of *Assessment in Education* over the last ten years is quite simply how international the field has become. Not only are new assessment policies and practices rapidly exported around the world, an increasing volume of assessment activity is explicitly international in being designed to compare national indicators and performance. In this section we briefly review some illustrative examples in this respect from past issues of *Assessment in Education* to highlight three key themes: first, the increasing willingness of researchers to acknowledge—and indeed to study—the impact of context in the operation of particular assessment practices; second, the global scale and impact of assessment policy and practice; and third, the development of international surveys of learning and achievement.

Clearly, ‘context’ can be defined in a variety of ways. In some cases the contexts in question are generic—such as classrooms or workplaces—but in many others it is a particular national context that provides the backdrop. The desire to understand the impact of context underpins the adoption of an explicitly comparative approach between two or more national settings, such as the study by Bray (1998) of examination practices in small states, or Sadler (1994). Other studies are single-country case studies of more international phenomena, such as Sebatane’s (1994) study of developing schoolteachers’ skills in classroom assessment in Lesotho, and Baumgart and Halse’s (1999) exploration of the impact of culture on student performance.

Specially commissioned ‘country profiles’ have provided such contextual insights for China, France, Spain, Lesotho, Nepal, Egypt, Singapore, Saudi Arabia, New Zealand and South Africa (see Appendix 2). Although the primary purpose of these country profiles is to be useful in themselves for readers anxious to know about the assessment practices of a particular country, their scope in this respect is inevitably limited in terms of their coverage of the world as a whole. However, the compensation is the basis they provide for more systematic comparative studies of assessment practices across countries that may share significant commonalities of history or culture—such as Australia and New Zealand, Hong Kong, China and Singapore, or France and Germany.

Increasing globalization inevitably means that different parts of the world are increasingly grappling with similar assessment challenges. One such is the vexed issue of selection for higher education explored in the special issue on entry into higher education edited by Bakker and Wolf (Vol. 8, No. 3) which draws on studies from a range of different countries to identify how different national education systems are seeking to resolve the more or less contradictory pressures of increasing enrolment whilst sustaining standards and public confidence. This theme is taken up more generally in the special issue ‘Globalization, qualifications and livelihoods’, edited by Little (Vol. 7, No. 3), which documents the impact of the global trade in qualifications and the significance of international economic trends for different countries, including some of the poorest and most vulnerable individuals and nations.

The global scale of assessment activity and the significance of this for all societies, rich and poor, was brought out powerfully by Dore's classic 1976 study, *The diploma disease*, which traced how the interaction between the supply of qualifications and the availability of modern sector jobs tended to result in the relentless pursuit of ever higher levels of 'diploma' as a form of educational inflation. Little's special issue, 'The diploma disease revisited' (Vol. 4, No. 1); provides a fascinating insight into the subsequent interaction of time, policy and economic developments in shaping the different trajectories of Dore's original assessment phenomenon in individual countries. It reveals increasing evidence of the powerlessness of individual countries to stand outside the now international market for qualifications. The range of developed and developing countries that form the empirical bases for the various contributions to these special issues provide powerful testimony to the need to understand assessment in relation to the particular political and policy context that is unique to that national setting.

The third and extremely important aspect of the increasing internationalization of assessment is that of the large-scale international assessment studies that have characterized recent decades. The defining role played by formal assessment procedures in determining individual futures is now increasingly being complemented by the rise of policies concerned with other, complementary, uses of assessment at institutional and system level. The assumptions of scientific measurement that inform the principles identified above are being used more and more to justify the application of measures to judge the relative achievements of institutions in the context of comparative league tables. Students' performance in public examinations and on specially set national tests has now become the legitimate currency for judgements of the quality of the educational process itself, as well as of individual merit.

These more recent international developments have been reflected in the pages of *Assessment in Education*. One strand of research has concerned international surveys of student achievement. The many technical pitfalls that surround attempts to compare national educational performances in different subjects were comprehensively explored in the 1996 special issue on the International Association for the Evaluation of Educational Achievement (IEA) studies, edited by Goldstein (Vol. 3, No. 2). Subsequent articles (e.g. Blum *et al.*, 2001) have developed these debates in the context of more recent international studies.

Several articles in the journal have brought out the many pitfalls in using large-scale test programmes, composed and interpreted outside the school sites where they are administered, in order to audit schools' performances and those of their students. Analyses of international studies have helped to draw attention to the many features outside the control of schools, which influence performance, and call into question the crude league table approach to the interpretation of performance results, either within or between countries. In showing that within-country variation is a greater source of variance than between-country variation Shen and Pedulla (2000) and Shen (2002) drew attention to the potential value of within-country analyses of survey results. At an even finer level of detail, the finding for Israel that within-school differences are more significant than between-school differences,

indicates that policy decisions taken on a macro level analysis of results may be ill-conceived (Yair, 1997). There is a contrast here between the potential for interpretation of international studies and the relative weakness, because of lack of rich background data, of the blanket testing exercises of which governments are increasingly enamoured.

However, large-scale test data can be called into question at a deeper level. The difficulties of allowing for variable opportunities to learn in interpreting tests designed to span the many variations between the curricula of different countries are notorious, and again draw attention to the need for interpretation at micro as well as at macro levels. Sample surveys designed for use within one country can explore a far wider range of test instruments and contexts outside the constraints of international surveys. This was illustrated by Murphy's (1995) use of Assessment of Performance Unit (APU) and National Assessment of Educational Progress (NAEP) data to explore gender bias, but she has also argued (Murphy, 1996) that the IEA tests would have to range over far more assessment instruments before they could be capable of valid interpretations. Cresswell (1994) reached a similar conclusion in his analysis of the different possible methods of aggregating and combining data for UK public examinations: the basic problem here is the attempt to make inferences on an inadequate set of data.

Cutting across the three themes discussed above is the importance of studying the impact on policy of the trends that they illustrate (Brown, 1996; Kellaghan, 1996). Governments have been quick to seize on the headline findings whether positive or negative, with little regard to the dependability of the data produced or the range of possible explanations for them. All too often, responses have been context-blind, the apparent successes of other countries being attributed to specifics of curriculum design or pedagogic strategies rather than the more subtle influences of culture and tradition. Optimistically, one could hope that by using their international leverage, researchers can encourage countries to invest effort more wisely to obtain information that can profitably guide policy.

Overall, however, a significant result of such research studies has been the creation of a world trade in educational policies, especially with regard to assessment. The phenomenon of 'policy-borrowing'—the take-up of apparently good ideas developed in one country by another—has further strengthened the grip of conventional assessment assumptions. Despite the significant evidence concerning flaws in international comparisons of student achievement, the power of the simple messages that can be and are derived from them about relative national success in a world of increasingly global competition has served significantly to reinforce the prevailing domination of established forms of educational assessment.

In addition to such explicitly international issues, *Assessment in Education* has carried many articles that offer case studies of particular national practices. Whether this concerns the way in which teachers are assessed for allocation to posts in France (Beattie, 1996) for example, or the teaching of maths in Europe, the journal offers a rich variety of material in which the reader is invited to take into account the particular 'package' of policy and practice that characterizes that setting.

Few indeed are the articles that are 'context-blind' in not recognizing—to a

greater or lesser extent—that assessment is a function of, and in turn impacts upon, particular educational settings. To the extent that this is so, it prompts the important question as to how far it is therefore legitimate to generalize about assessment phenomena across different social settings. To what extent can a journal that purports to be interested in identifying internationally valid insights to inform and guide the future development of assessment policy and practice offer generalizations that do not ride roughshod over the significant effects of context? In short, how can we distinguish the ‘constants’ from the ‘contexts’ in assessment practice? (Blum *et al.*, 2001; Broadfoot, 2001).

The answer to this question must lie in understanding the commonalities that have informed the historical development of assessment around the world and hence, some of the challenges that countries are likely to share in addressing the changes needed to make future assessment principles, policies and practices fit for purpose. It is rare for assessment scholars to look backwards. The millennium issue of *Assessment in Education*, edited by Harvey Goldstein and Gillian Sutherland (Vol. 8, No.1), offers a rare example in this respect. Its theme of ‘Past, present and future’ embraces contributions that range from assessment in ancient Rome (Morgan, 2001) to the potential impact of globalization in the twenty-first century (Kellaghan & Greaney, 2001).

### Quality concerns

The quality of any particular assessment is typically addressed in terms of measures of reliability and validity. One of the yawning gaps in public test information in many countries is the lack of any comprehensively researched data on their reliability, even although the technical means for doing this are well known. The studies by Rogosa (1999) in California, and by Gardner and Cowan (2000) in Northern Ireland, are among the few available studies to address this issue. They show how alarmingly large are the chances of students being wrongly graded, and in the latter the authors draw attention to the fact that tests which are nationally important and can determine life chances of children do not satisfy the standards for testing of the professional test community (AERA/APA/NCME, 1999).

Whilst there has been little attention to this issue of reliability in the pages of *Assessment in Education*, there have been significant contributions to the literature on validity. In an important paper in Vol. 3, Crooks *et al.* (1996) used the metaphor of a chain to indicate the linked set of factors that can limit validity, raising the possibility of improvement by targeting the weakest link. Their choice of weakest link was the interface of question with student, where student anxiety, motivation, understanding of the process and of the language used, and the whole context of the encounter can all affect performance in ways for which users remote from that context cannot make allowances. Gipps (1995) spelt out a similar argument, stressing that only with multimodal tasks encountered in non-threatening settings could unintended bias between different students and schools be reduced. A different problem in validity was raised by Husén and Postlethwaite (1996), querying whether precise interpretations of a learner’s intellectual functioning can be

inferred from limited test responses. There is a pressing need for more statistical analyses of multiple sources of data in this respect, in order to infer probabilities for various inferences.

The most comprehensive view of validity was that expressed by Messick (1989) who argued that the validity of assessment practices inheres first and foremost in the consequences that follow from their use. The impact of assessment on the lives of individuals is becoming more widespread and serious with its growing importance across the world (a growth amply illustrated in Vol. 7, No. 3). Thus it is arguably as strong a moral imperative on educators to satisfy themselves that the technologies being used are 'safe' and beneficial as there is on a nuclear scientist or on a biologist working on genetically modified crops.

It follows that there is clearly a need for more thorough explorations of both the validity and the reliability of the various approaches to designing and interpreting the test data which are commonly used by governments and which command the confidence of a public which does not understand the technical limitations. The research data show that current policies are ill-informed, and are almost certainly far from optimum, though rich and varied. Studies that synthesize these in order to define, look for and evaluate potentially more robust approaches are urgently needed.

### **Assessment for learning**

The value of formative assessment practices was firmly established in Vol. 5, No. 1, where the review by Black and Wiliam (1998) was complemented by a range of respondent articles. The significant impact of that review, notably on some subsequent policy shifts in the UK, but also in supporting other work on formative assessment in other countries, owes much to its emphasis on the warrants for the claims of such work provided by the quantitative evidence of learning gains. A significant addition to this evidence was the article by McDonald and Boud (2003) showing strong gains in examination scores when pupils are trained in self-assessment and peer assessment; however, it is noteworthy that this is the only paper in the journal that presents a classic quantitative experiment-control test of a hypothesis.

*Assessment in Education* has undoubtedly played a leading part in raising the debate about assessment in the service of learning, with over 40 papers focusing on how this is implemented in classroom assessment. In diverse ways these enrich our understanding of the complexity of the links between assessment and learning, and in so doing provide much evidence to inform what is now an emerging issue—the link between teachers' practices in formative and in summative assessment, and the prospects for strengthening the quality and status of teachers' summative assessments. The issue is important in two ways. The first is that, as formative assessment becomes more clearly recognized and is implemented in teachers' practices, conflicts with the requirements of concurrent summative assessments are bound to inhibit and even frustrate. The second is that one way to overcome the severe limitations of

external testing clearly must be to use teachers' own knowledge of their students as a source of data for the purposes of certification and accountability.

Several studies (e.g. Morgan, 1996; Preece & Skinner; 1999; Shen, 2002) have shown how summative test requirements dominate the assessment practice of many teachers. The dominance is not merely a product of external test requirements: in the context of USA primary classrooms Bachor and Anderson (1994) found that teachers do not distinguish between formative and summative purposes, whilst Mavrommatis (1997) found that in primary classrooms in Greece teachers resisted attempts at reform which would challenge their summative habits, even to the extent of continuing to give grades on written work after this had been declared illegal. The fine-grained analysis of Torrance and Pryor (1995) explored the complex of factors that bear on teachers' practice, leading much of that practice to be closer to frequent summative rather than formative assessment. This study makes particularly clear that such detailed research on daily practice may be necessary if programmes of change are to be firmly grounded. A comparable study by Cowie and Bell (1999) proposed a model which distinguishes between planned and interactive formative assessment; the latter is more demanding, and its practice is more fragile under stress.

If formative assessment is to prosper, initiatives aimed at supporting a positive link between formative and summative work are sorely needed. The common model in the UK, for school-leaving examinations, is to require teacher assessment of set pieces of students' work which will cover aspects of the curriculum, notably practical work, which written tests cannot explore. This approach can have a deplorable effect: the early paper by Paechter (1995) exposed how, as UK teachers had to 'administer' tasks constrained by rules, teachers were uncertain of their role, some behaving as external examiners, others refusing to suspend the normal teaching role which they would play with such tasks. Similar pictures of ambiguity and tension were described, in the context of US innovations, under the broad umbrella title of 'performance assessment', by Baker and O'Neill (1994).

A different approach is portfolio assessment, which was featured in a special issue in 1998 (Vol. 5, No. 3). The papers brought out a sharp contrast between the attractions of the freedom this approach gave to teachers and their students, and the weak features that have all but derailed the initiative. One aspect was brought out by Stecher (1998) showing how teachers' practices were narrowed down to 'rubric-driven instruction' as requirements of reliability and validity imposed constraints. Koretz (1998) documented the disappointments in three different initiatives in the USA: weak reliability and validity followed from the lack of adequate training to help teachers be consistent and rigorous in the framing, the selection, and the assessment of portfolio components. This is a clear case of the clash between potential learning advantages and social beliefs about what is required from schools—but it is not possible to say whether both can be satisfied. It is even possible that professional development meant to increase the consistency of assessment could actually improve the quality of the support for student learning.

A more positive prospect for both enriching and underpinning teachers' summative work can be envisaged by external provision of test instruments for teachers to

use at their discretion. If they are to be useful for learning, such instruments should be developed to reflect progression in a subject—a requirement which calls for rigorous attention to issues of equity and construct validity in the framing of questions, and empirical development to ensure scalability, that is, that the sequence of items is affirmed by the sequence in which students actually succeed. Such rigour can call into question the conceptual consistency of existing standards—or, in UK language, levels of attainment—as shown in the Davies (2002) study of these levels in relation to learning development in geography. This is not surprising, given that government prescriptions have usually been formulated without evidence of ‘scalability’. Both Gilbert (1996) in the UK and Rowe and Hill (1996) for Australia describe the provision of well-researched resources, with Gilbert stressing that the development in art has promoted valuable ‘assessment conversations’ both between teachers, and between teachers and their students.

More radically, recent papers looking to the needs, aspirations and perceptions of students call for new thinking. The need to attend to the students’ perspective has played too small a part in research on assessment. Notable exceptions in this journal are the papers by Moni *et al.* (2002) and by Brookhart and Bronowicz (2003), which show that students may well interpret all assessments as summative, and both devalue and/or resist their involvement in them. The papers by Newfield *et al.* (2003) and by Johnson (2003) draw attention to the needs of disadvantaged urban youth in developing countries. Both emphasize that such students can only achieve if assessments allow them to use those modes of expression in which they can be fluent, and this means that multimodal expressions, involving varied combinations of art, craft, oral and verbal modes, must be recognized. In a different context, Jewitt (2003) argued that ICT-based work produces outcomes in a mode that current criteria do not recognize. Any such radical developments raise questions about who promulgates criteria and whose interests they serve, and so brings the argument round to the context of social control within which all assessment activity is framed.

None of this leads to recipes for aligning formative and summative practices—there is far more work to be done if the optimum synergy between these two, and so between assessment for learning and assessment for certification and accountability, is to be achieved.

### **The way forward: redefining assessment?**

Much of this article has looked back, as is appropriate for an anniversary issue. In it we have tried to highlight some of the more significant debates that have characterized international assessment research in recent years as these have been represented in the pages of *Assessment in Education*. We have structured these reflections in terms of four key topics that seem to us to have been the focus of particular attention during this period: the purpose of assessment; international perspectives; issues of quality; and assessment for learning. We took as a starting point the perspective of the journal as a whole on *principles, policy and practice* in assessment, which highlight the spectrum of influences and contexts that shape the social realization of assessment. It is appropriate in this final section that we try to

draw together these threads in order to provide some kind of overview of these more detailed analyses, and hence, to delineate what has been achieved in the first ten years of the journal's existence. It is also appropriate that we look ahead to the next ten years or more to try and discern what are likely to be the key assessment topics and tensions over this period and perhaps even to influence this agenda in some small way.

Some of the defining aspects of recent assessment research stand out with quite remarkable clarity. Chief amongst these is the increase in assessment activity of all kinds and the penetration of assessment in its various guises into almost every aspect of human endeavour. In very truth, we have become an 'assessment society', as wedded to our belief in the power of numbers, grades, targets and league tables to deliver quality and accountability, equality and defensibility as we are to modernism itself. History will readily dub the 1990s when this journal was born—as well as the early years of the new millennium—'the assessment era', when belief in the power of assessment to provide a rational, efficient and publicly acceptable mechanism of judgement and control reached its high point. It is probably no accident that this development came at a time when capitalism itself became transformed into a global system and the other trappings of globalization—instant international communication and the knowledge economy—also developed in previously almost unimaginable ways, a decade during which email and the World Wide Web for example, have transformed all our lives.

The assessment revolution—as it may properly be called—has been one of scale, range and significance; a revolution that has elevated *quantitative data*—the raw material of most public assessment—as the principal mechanism for delivering transparency, accountability and predictability. The collection of data has become in itself a major instrument of social control, whether this is at the level of the individual, the institution or indeed whole operational systems such as that of education. The impact of decisions taken concerning what data to collect and how, are likely to have a significance and an effect far beyond the task in hand in what has become widely known as the 'wash-back' effect (Messick, 1989).

This growth in assessment activity as a political and policy tool and its consequent prominence as a public issue has helped to underpin the success of a journal in which the focus is as much social as technical. The journal has in turn helped to develop the emerging challenge to the dominance of the prevailing assessment Zeitgeist. The reification of learning outcomes, the extreme rationalism of codes for quality which attempt to rank and compare on the basis of reductionist measures, the decontextualization of national educational performance from issues of culture and context, are all reflections of this Zeitgeist; of a modernist view of the legitimate use of assessment data to name, to compare and to judge.

In recent years, however, it is arguably the very pervasiveness of such approaches that has fuelled a growing challenge to received orthodoxies. The worldwide tendency for more young people to stay on longer in formal education that now increasingly includes higher education, coupled with a growing discourse of 'lifelong learning', has helped to shift attention towards how best to support students' learning, rather than to judge it. The well-established technical limitations of

conventional tests and examinations are increasingly becoming an issue in a world of high-stakes tests and growing litigation. An emerging concern with ethical and human rights issues is also beginning to fuel an examination of the principles that should inform assessment. In some countries it is the sheer volume of assessment that young people are now exposed to in school, coupled with the growing inability of existing assessment technologies to cope with the increased frequency and scale of formal testing, that is beginning to prompt a radical rethink of both what is possible and what is desirable. As professionals themselves increasingly come to feel the heavy hand of assessment and the tensions and anxieties associated with it, perhaps they are becoming more willing to empathize with the extraordinary diet, in many countries, of relentless judgements to which it now seems necessary to expose young people.

All these various straws in the wind are helping to challenge the assumptions on which most of the existing edifice of assessment has been built. Belief in the power of conventional summative assessment techniques to be objective and efficient, to motivate present performance and to predict future performance, is being challenged by a range of research evidence that identifies the significant flaws in these assumptions. Moreover, the assumptions highlight the worrying price that the use of assessment to measure and control extracts, including reduced motivation and significantly lower performance on the part of students (Black & William, 1998; Harlen & Deakin-Crick, 2003). At a more macro level 'jumping through assessment hoops' and 'playing the league tables game' are not inappropriate metaphors for what is increasingly seen as a poor substitute for genuinely enhancing the quality of delivery systems in the ways required according to the arguments outlined in our 'Quality concerns' section above.

Much of the familiar contemporary apparatus of assessment technologies was born of the modernist assumptions and educational needs of the nineteenth century (Madaus *et al.*, 1997; Broadfoot, 1998; Thomas *et al.*, 1998). The assumptions informing these approaches may be identified as:

- (1) That it is right, 'objectively' to seek to identify relative levels of student performance as the basis for educational selection.
- (2) That it is possible to undertake such identification with a sufficient degree of 'objectivity' that it provides a broadly fair outcome for the candidates affected.
- (3) That the quality of such assessment is embodied in notions of reliability and validity.
- (4) That students' scores on national examinations and tests provide a valid indicator of the quality of institutional performance.
- (5) That it is possible usefully to compare the 'productivity' of individual education systems through international comparisons.

So taken for granted are such assumptions that they are rarely articulated in this kind of way. By virtue of their ubiquitous presence in contemporary education systems, the necessity for, and the desirability of, educational assessment in these terms is rarely, if ever, questioned. It is simply impossible to imagine formal education without the periodic punctuation of assessment events designed to check

up, report on, certificate or select individual students. Just occasionally such assumptions—and associated definitions of good practice—are more explicitly articulated. Two examples which were published in *Assessment in Education* are: the ‘Principles for fair student assessment practices for education in Canada’ (Rogers, 1996) and the ‘Code of professional responsibilities in educational measurement’ (Vol. 3, No. 3, pp. 401–411). The manual for testing standards produced jointly in the USA by the leading research bodies (AERA/APA/NCME, 1999) is another example.

Welcome as such initiatives are, they are rare indeed in comparison to the scale, cost and impact of educational assessment practices. A more concerted international articulation of defensible principles would be highly desirable, especially since such an articulation would be likely to challenge developers, users, critics and the general public to hold such practices up to an unprecedented degree of scrutiny in terms of their fitness for purpose. It may be, as those professionally involved in the development and delivery of assessment systems have long accepted, that the inevitable limitations of attempting to squeeze human variability into a scientific discourse results in what is at best a rough and ready system, but one that is acceptable as the least worst option for meeting pressing educational and social needs.

The scale of the growth in assessment as an international policy tool reinforces the grave responsibility of the community of assessment scholars worldwide to be mindful of the influence their work is likely to have on many millions of students and teachers world-wide. It is clear that the assessment community must strive to be as clear ethically as it is technically about the appropriate uses of the results of their research. Arguably, one of the great contributions of *Assessment in Education* has been to bring a long overdue and very necessary discussion of purposes and consequences more to the fore in assessment research.

It is likely that a thorough-going cost–benefit analysis of existing assessment practices would reveal a significant mismatch between the principles on which current forms of assessment are largely based and the capacity of available techniques adequately to meet these principles. Even more seriously, it might articulate a different set of assessment principles born of educational, rather than measurement, priorities, which could in turn call into question the legitimacy of much contemporary assessment thinking.

In particular, we suggest, the following questions are in urgent need of attention:

- How far do prevailing modes of student assessment tend to reinforce outmoded notions of curriculum content and student learning at the expense of twenty-first century learning skills and dispositions such as creativity and learning to learn?
- How far are the comparisons of institutional and system performance that are currently being made defensible in terms of the accepted methodologies of comparative social research?
- To what extent has research into educational assessment focused to an excessive extent on techniques, at the expense of more fundamental analyses of its social and political role, and so failed to mount a concerted challenge to the inappropriate and damaging use of such techniques?

- Is it now time for the emergence of a new assessment paradigm born of the very different epistemologies and needs of the twenty-first century?

One of the most encouraging insights to have emerged from the pages of *Assessment in Education* in recent years, as well as from the world of assessment research more generally, is that assessment can be a powerful force in supporting learning, and a mechanism for individual empowerment. It can help learners at all ages and stages to become more self-aware, more expert in mapping an individual learning path in relation to their own strengths and weaknesses and in facilitating fruitful collaboration with fellow learners.

As the millennium issue of *Assessment in Education* so clearly illustrated in linking past and future perspectives, at various times in history educational assessment has played a powerful role in facilitating social progress. Thus, just as, for example, the advent of formal examinations in nineteenth-century Europe arguably represented a major victory for social justice, so the advent of a focus on assessment for learning in the twenty-first century could be similarly significant for social reform. Clearly, any such trend is likely to be part of much broader social and ideological developments. The shift towards what Habermas has termed 'emancipatory rationality', with its focus on the human potential for self-realization and creativity rather than on the hermeneutic rationality of data, systems and bureaucracy, encapsulates both the nature of the challenge that confronts us and the potential reward that awaits its successful resolution. In a world in which human beings find themselves increasingly cut off from well-defined norms, community support and collective goals, it becomes increasingly necessary to find ways of helping them to be able to define themselves as individuals and to cope with managing their own learning and work careers.

A significant volume of assessment research in recent years has been instrumental in shaping understanding about what might be done in this respect and how. Portfolios and records of achievement, self-assessment and action planning are all relatively recent ideas that reflect a very different role for assessment. Recent developments around assessment for learning represent more explicitly curriculum-focused aspects of the same agenda. But the scope is potentially much greater than this and it seems likely that we are only as yet aware of the tip of the iceberg of what a new assessment paradigm could be like. In counselling, for example, the notion of 'restorative' assessment has been developed, in which assessment techniques are explicitly used to help repair damage caused by the pressures of contemporary lifestyles (Speedy *et al.*, 2003).

Thus it seems likely the next ten years of *Assessment in Education* will document and, hopefully, contribute to major new developments in assessment thinking and research. We have argued here that the next few years may see a shift in the 'normal' science of assessment (Kuhn, 1962) on a scale that is, as yet, hard to imagine. If this does indeed prove to be the case, it will be a tribute to the efforts of the assessment research community around the world, in the light of an emerging discourse around assessment principles, to examine critically and systematically the panoply of assessment policies and practices that have evolved during the preceding century.

Assessment technologies are just as powerful in their own way as others such as IT and genetic modification. Yet with regard to assessment, this power to help or to harm has gone much less remarked upon in comparison with the enormous use made of the technologies. If *Assessment in Education* has gone some way to rectifying this situation; if it contributes to changing for the better, even in a small way, the lives of young people around the world; if it helps to moderate the tidal wave of targets and accountability and instead is employed in liberating the power of learning to change lives, this will be a fitting tribute to its founding vision and to the first ten years of its life.

### Notes on contributors

Patricia Broadfoot is Professor of Education and Pro Vice-Chancellor of the University of Bristol. She was a founding editor of the Journal and managing editor from its inception until 2002. She has written widely in the field of assessment from a sociological perspective. Her most recent book, *Assessment: what's in it for schools?* with Paul Weeden and Jan Winter (Routledge/Falmer, 2002) explores how often assessment fails to achieve its intended purpose and how teachers can be helped to harness it more effectively to support students' learning.

Paul Black is Emeritus Professor at King's College London. During his career he has been involved in a range of Nuffield curriculum projects and in many research projects, mainly in science education and assessment. In 1987–8 he chaired the task group (TGAT) that advised ministers on the new national assessment and testing policy. Since his retirement he has concentrated on the study of formative assessment.

### Note

The authors would like to acknowledge the significant contribution to the founding of *Assessment in Education* of the late Professor Desmond Nuttall, whose seminal contributions to assessment research in the cause of Improving the Quality of Education are celebrated in an early special issue of the journal (Vol. 1, No. 3).

### References

- Airasian, P. W. & Gregory, K. D. (1997) The Education Reform Act of 1998 (Review essay), *Assessment in Education*, 4(2), 307–314.
- American Educational Research Association (AERA)/American Psychological Association (APA)/National Council on Measurement in Education (NCME) (1999) *Standards for educational and psychological testing* (Washington DC, AERA)
- Bachor, D. G. & Anderson, J. O. (1994) Elementary teachers' assessment practices as observed in the province of British Columbia, Canada, *Assessment in Education*, 1(1), 63–94.
- Baker, E. L. & O'Neil, H. F. (1994) Performance assessment and equity: a view from the USA, *Assessment in Education*, 1(1), 11–26.
- Baumgart, N. & Halse, C. (1999) Approaches to learning across cultures: the role of assessment, *Assessment in Education*, 6(3), 321–339.
- Beattie, N. (1996) Interview and *concours*: teacher appointment procedures in England and Wales and France, and what they mean, *Assessment in Education*, 3(1), 9–28.

- Black, P. & Wiliam, D. (1998) Assessment and classroom learning, *Assessment in Education*, 5(1), 1–74.
- Blum, A., Goldstein, H. & Guérin-Pace, F. (2001), International Adult Literacy Survey (IALS): an analysis of international comparisons of adult literacy, *Assessment in Education*, 8(2), 225–246.
- Bray, M. (1998) National self-determination and international dependence: the organisation and control of secondary school examinations in the small states of the Commonwealth, *Assessment in Education*, 5(2), 151–173.
- Broadfoot, P. (1998) Records of achievement and the learning society: a tale of two discourses, *Assessment in Education*, 5(3), 447–477.
- Broadfoot, P. (2001) Editorial: new wine in old bottles? The challenge of change for educational assessment, *Assessment in Education*, 8(2), 109–112.
- Brookhart, S. M. & Bronowicz, D. L. (2003) ‘I don’t like writing. It makes my fingers hurt’: students talk about their classroom assessments, *Assessment in Education*, 10(2), 221–242.
- Brown, M. (1996) FIMS and SIMS: the first two IEA International Mathematics Surveys, *Assessment in Education*, 3(2), 193–212.
- Cowie, B. & Bell, B. (1999) A model of formative assessment in science education, *Assessment in Education*, 6(1), 101–116.
- Cresswell, M. J. (1994) Aggregation and awarding methods for National Curriculum assessments in England and Wales: a comparison of approaches proposed for Key Stages 3 and 4, *Assessment in Education*, 3(1), 45–61.
- Crooks, T., Kane, M. T. & Cohen, A. S. (1996) Threats to the valid use of assessment, *Assessment in Education*, 3(3), 265–286.
- Davies, P. (2002) Levels of attainment in geography, *Assessment in Education*, 9(2), 185–204.
- Dore, R. (1976) *The diploma disease* (London, Unwin).
- Eckstein, M. A. & Noah, H. J. (1993) *Secondary school examinations: international perspectives on policies and practice* (New Haven, CT, Yale University Press).
- Firestone, W. A. (1998) A tale of two tests: tensions in assessment policy, *Assessment in Education*, 5(2), 175–191.
- Gardner, J. & Cowan, P. (2000) *Testing the test; a study of the reliability and validity of the Northern Ireland transfer procedure test in enabling the selection of pupils for grammar school places* (Belfast, Queen’s University of Belfast).
- Gilbert, J. (1996) Developing an assessment stance in primary art education in England, *Assessment in Education*, 3(1), 55–74.
- Gipps, C. (1995) What do we mean by equity in relation to assessment? *Assessment in Education*, 2(3), 271–282.
- Harlen, W. & Deakin-Crick, R. (2003) Testing and motivation for learning, *Assessment in Education*, 10(2), 169–208.
- Harlen, W. & James, M. (1997) Assessment and learning: differences and relationships between formative and summative assessment, *Assessment in Education*, 4(3), 365–379.
- Husén, T. & Postlethwaite, T. N. (1996) A brief history of the International Association for the Evaluation of Educational Achievement (IEA), *Assessment in Education*, 3(2), 129–142.
- Jewitt, C. (2003) Re-thinking assessment: multimodality, literacy and computer-mediated learning, *Assessment in Education*, 10(1), 83–102.
- Johnson, D. (2003) Activity theory, mediated action and literacy: assessing how children make meaning in multiple modes, *Assessment in Education*, 10(1), 103–129.
- Kellaghan, T. (1996) IEA studies and educational policy, *Assessment in Education*, 3(2), 143–160.
- Kellaghan, T. & Greaney, V. (2001) The globalisation of assessment in the twentieth century, *Assessment in Education*, 8(1), 87–102.
- Klenowski, V. (1995) Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England, *Assessment in Education*, 2(2), 145–164.
- Koretz, D. (1998) Large-scale portfolio assessments in the US: evidence pertaining to the quality of measurement, *Assessment in Education*, 5(3), 309–334.

- Kuhn, T. (1962) *The structure of scientific revolutions* (Chicago, IL, University of Chicago Press).
- Madaus, G. F., Clarke, M. & O'Leary, M. (1997) The kaleidoscope of assessment: disciplinary angles (Review essay), *Assessment in Education*, 4(3), 431–438.
- Mavrommatis, Y. (1997) Understanding assessment in the classroom: phases of the assessment process—the assessment episode, *Assessment in Education*, 4(3), 381–400.
- McDonald, B. & Boud, D. (2003) The impact of self assessment on achievement: the effects of self assessment training on performance in external examinations, *Assessment in Education*, 10(2), 209–220.
- Messick, S. (1989) Validity, in: R. L. Linn (Ed.) *Educational measurement* (3rd edn) (London, Collier Macmillan), 12–103.
- Moni, K. B., van Kraayenoord, C. & Baker, C. D. (2002) Students' perceptions of literacy assessment, *Assessment in Education*, 9(3), 319–342.
- Morgan, C. (1996) The teacher as examiner: the case of mathematics coursework, *Assessment in Education*, 3(3), 353–376.
- Morgan, T. (2001) Assessment in Roman education, *Assessment in Education*, 8(1), 11–24.
- Morrison, K. & Tang Fun Hei, J. (2002) Testing to destruction: a problem in a small state, *Assessment in Education*, 9(3), 289–317.
- Murphy, P. (1995) Sources of inequity: understanding students' responses to assessment, *Assessment in Education*, 2(3), 249–270.
- Murphy, P. (1996) The IEA assessment of science achievement, *Assessment in Education*, 3(2), 213–232.
- National Council on Measurement in Education (NCME) (1995) *Code of professional responsibilities in educational assessment*. Reproduced in *Assessment in Education*, 3(3), 401–411.
- Newfield, D., Andrew, D., Stein, P. & Maungedzo, R. (2003) 'No number can describe how good it was': assessment issues in the multi-modal classroom, *Assessment in Education*, 10(1), 61–81.
- Paechter, C. (1995) 'Doing the best for the students': dilemmas and decisions in carrying out statutory assessment tasks, *Assessment in Education*, 2(1), 39–52.
- Preece, P. F. W. & Skinner, M. C. (1999) The national assessment in science at Key Stage 3 in England and Wales and its impact on teaching and learning, *Assessment in Education*, 6(1), 11–26.
- Rogers, T. W. (1996) Principles for fair student assessment practices for education in Canada, *Assessment in Education*, 3(3), 397–400.
- Rogosa, D. (1999) *How accurate are the STAR national percentile rank scores for individual students?—an interpretive guide*. CSE Technical Report 509a (Los Angeles, CA, CRESST). Available online at: [http://www.cse.ucla.edu/products/reports\\_set.htm](http://www.cse.ucla.edu/products/reports_set.htm)
- Rowe, J. R. & Hill, P. W. (1996) Assessing, recording and reporting students' educational progress: the case for 'subject profiles', *Assessment in Education*, 3(3), 309–352.
- Sadler, R. (1994) Examinations and merit (Review of *Secondary school examinations* by M. A. Eckstein & H. J. Noah), *Assessment in Education*, 1, 115–120.
- Sebatane, E. M. (1994) Enhancement of teacher capacities and capabilities in school-based assessment: Lesotho experience, *Assessment in Education*, 1(2), 223–234.
- Shen, C. (2002) Revisiting the relationship between students' achievement and their self-perceptions: a cross-national analysis based on TIMSS 1999 data, *Assessment in Education*, 9(2), 161–184.
- Shen, C. & Pedulla, J. J. (2000) The relationships between students' achievement and self-perception of competence and rigour of mathematics and science, *Assessment in Education*, 7(2), 237–254.
- Speedy, J., Winter, J., Broadfoot, P., Thomas, J. & Cooper, B. (2003) Researching assessment cultures, re-researching ourselves, in: R. Sutherland, G. Claxton & A. Pollard (Eds) *Learning and teaching where worldviews meet* (Stoke-on-Trent, Trentham Books), 255–271.
- Stecher, B. (1998) The local benefits and burdens of large-scale portfolio assessment, *Assessment in Education*, 5(3), 335–352.

- Thomas, S., Madaus, G. F., Raczek, A. E. & Smees, R. (1998) Comparing teachers' assessment and standard task results in England: the relationship between pupils' characteristics and attainment, *Assessment in Education*, 5(2), 213–246.
- Torrance, H. & Pryor, J. (1995) Investigating teacher assessment in infant classrooms: methodological problems and emerging issues, *Assessment in Education*, 2(3), 305–320.
- Yair, G. (1997) When classrooms matter: implications of between-classroom variability for educational policy in Israel, *Assessment in Education*, 4(2), 225–248.

## **Appendix 1. *Assessment in Education special issues, 1994–2003***

Vol. 1, No. 3 (November 1994)

*Improving the Quality of Education: a Memorial Issue in Honour of the Work of Professor Desmond Nuttall*

Patricia Broadfoot (Ed.)

Vol. 3, No. 2 (July 1996)

*The IEA Studies*

Harvey Goldstein (Ed.)

Vol. 4, No. 1 (January 1997)

*The Diploma Disease Twenty Years On*

Angela Little (Ed.)

Vol. 5, No. 1 (March 1998)

*Assessment and Classroom Learning*

Patricia Broadfoot (Ed.)

Vol. 5, No. 3 (November 1998)

*Portfolios and Records of Achievement*

Dan Koretz, Patricia Broadfoot and Alison Wolf (Eds)

Vol. 6, No. 1 (March 1999)

*Science Education*

Wynne Harlen (Ed.)

Vol. 7, No. 3 (November 2000)

*Globalisation, Qualifications and Livelihoods*

Angela Little (Ed.)

Vol. 8, No. 1 (March 2001)

*Assessment at the Millennium*

Harvey Goldstein and Gillian Sutherland (Eds)

Vol. 8, No. 3 (November 2001)

*Upper Secondary Examinations and Entry to University: the school-university transition in an age of mass higher education*

Steven Bakker and Alison Wolf (Eds)

Vol. 10, No. 1 (March 2003)

*Assessment, Literacies and Society: redesigning pedagogy and assessment*

David Johnson and Gunther Kress (Eds)

Vol. 10, No. 3 (November 2003)

*Assessment for the Digital Age*

Angela McFarlane (Ed.)

## **Appendix 2. Assessment in Education country profiles**

Vol. 3, No. 1 (March 1996)

*Educational Assessment in China*

Wang Gang

Vol. 4, No. 2 (July 1997)

*Country Profile from France*

Gerard Bonnet

Vol. 4, No. 3 (November 1997)

*Assessment in the Spanish Educational System*

Fernando Mahuenda

Vol. 5, No. 2 (July 1998)

*Educational Assessment in Lesotho*

E. Molapi Sebatane

Vol. 6, No. 1 (March 1999)

*Educational Assessment in Malta*

Ronald G. Sultana

Vol. 6, No. 3 (November 1999)

*Educational Assessment in Singapore*

Elaine P. Y. Lim and Annie Tan

*Public Examinations in Hong Kong*

Choi Chee-Cheong

Vol. 7, No. 1 (March 2000)

*Educational Assessment in Saudi Arabian Schools*

Ibrahim A. Al-Sadan

Vol. 7, No. 2 (July 2000)

*Assessment in Nepal*

Pramilla Rajbhandari and John Wilmot

Vol. 8, No. 2 (July 2001)

*Assessment in Egypt*

Eleanor Hargreaves

Vol. 9, No. 2

*Educational Assessment in New Zealand Schools*

Terry J. Crooks

*Assessment in South African Schools*

R. Cassius Lubisi and Roger J. L. Murphy

