



Bulletin number 19

Assessment: A Literature Review

Prepared for SQA by Mitch Miller June–July 2005

August 2006

Published by the Scottish Qualifications Authority
The Optima Building, 58 Robertson Street, Glasgow G2 8DQ

Ironmills Road, Dalkeith, Midlothian EH22 1LE

www.sqa.org.uk

The information in this publication may be reproduced in support of SQA qualifications. If it is reproduced, SQA should be clearly acknowledged as the source. If it is to be used for any other purpose, then written permission must be obtained from the Publishing Team at SQA. It must not be reproduced for trade or commercial purposes.

SQA is committed to the use of robust evidence in the development and evaluation of policy and its implementation, and carries out or commissions research across a range of topics to support this.

The publication of Research Bulletins allows us to disseminate the results of our research activity to practitioners, policy makers, parents, academics and anyone else who has an interest in the key role that qualifications play in economic growth and social inclusion in Scotland.

Contents

1	Introduction	1
2	Assessment theory and practice — general	7
3	Formative and diagnostic assessment	13
4	Summative assessment	17
5	Integrative assessment	20
6	Teacher assessment	21
7	Computer Assisted Assessment	24
8	Work-based assessment	28
9	Test/exam performance	32
10	Gender	34
11	Race and ethnicity in assessment	38
12	Disability	40
13	Comments and recommendations	41

1 Introduction

The following review was undertaken over June and July 2005. It outlines major debates in assessment research, practice and theory, with particular regard to formative, as well as summative, assessment and the impacts of social and personal factors. Such factors include age, race, gender and disability, the latter having stimulated many new developments in assessment practice and thinking to ensure that assessment methods do not unfairly disadvantage a candidate.

In short, we have come a long way since the days when examinations were the make or break for a student. As well as making a clear separation between the ability of a pupil in a subject and their ability to sit tests, current assessment theory has developed a distinct set of criteria and concepts to clarify the goals and aims of assessment in a range of different environments. The development of Computer Assisted Assessment, and its gradual implementation into educational systems, has ensured that assessment is highly topical in educational literature at the moment.

The review follows the 3–14 curriculum review, potentially the most significant recent shift in Scottish educational policy and practice, with many implications for assessment. As 3–14 continues to throw up new questions and challenges for SQA and the wider educational sector, it seems appropriate to ‘take stock’ of existing knowledge and current thinking on the topic. With this in mind, the purpose of this review was to:

- ◆ **Refresh thinking:** Is SQA’s use of criterion-referenced assessment a reflection of what works best, or simply tradition? Are there ideas from other educational systems that could be piloted here? How does Scotland compare, in its assessment regime, with other nations?
- ◆ **Inform future research:** Does the currently published research answer all of our questions about assessment? Are there any gaps in either general or specific knowledge that SQA or its associates might address?

The review begins with a summary of the literature studied and the main themes: History of Assessment, Assessment Theory and Practice, Summative Assessment, Formative and Diagnostic Assessment, Integrative/Synoptic Assessment and Environmental Factors (gender, race, disability). The commentary is followed by an annotated bibliography that lists the texts reviewed, and provides concise information about the most important of these alongside their details:

- ◆ Historical
- ◆ Formative and diagnostic
- ◆ Summative
- ◆ Integrative

- ◆ Teacher assessment
- ◆ Computer Assisted Assessment
- ◆ Work-based assessment
- ◆ Test/exam performance — general
- ◆ Gender
- ◆ Race/ethnic
- ◆ Disability

Although some work by authors based in, or associated with, SQA is included in this review, it is assumed that the purpose of this review is to look outwards — hence very few SQA documents are cited.

Methodology

This review was compiled partly through retracing the bibliographic record from earlier SQA assessment studies and obtaining new references. The review worked through the bibliographies of major reports and works, reference to major research centres and units, and advice from assessment experts within SQA. The internet was also used extensively, both to track down references cited in bibliographies and to acquire ‘e-texts’ and internet-only sources. The basic criteria applied in evaluating texts were:

- ◆ How relevant is it to the subject matter?
- ◆ How recent and up-to-date is it?
- ◆ To what degree is its text authoritative? Does it add a new dimension or avenue of debate/inquiry to the subject?
- ◆ How available is it? Is it still in print, available online or through library services?
- ◆ What are the credentials of the authors and/or research units involved?

An attempt was made whenever possible to consult the original text. Obviously, not even where the text could be secured was it possible to read the entire piece. A group of what appeared to be major texts was selected for reading, according to the following criteria (although they need not meet all of them):

- ◆ Major policy documents from Scottish, UK and European sources.
- ◆ A specifically Scottish focus or relevance.
- ◆ UK and international sources whose research had the widest application across political and cultural differences.
- ◆ Projects not yet completed that are likely to publish significant findings or recommendations in the future.
- ◆ The seeming ‘popularity’ of the text in citations and referencing, or a sense that the text is ‘seminal’ to the subject area.
- ◆ Unusual or unconventional takes on the subject.

- ◆ Indications of significant implications for policy and practice.

All texts that could be acquired were at least ‘skim-read’ and checked for conclusions and major points included in the synopses in the bibliographies. Throughout, the aim was to identify the shape and direction of debates and theories in assessment, and to secure those texts that seemed representative of these. The bibliographies are not by any means exhaustive, but do aim to collate and represent the most relevant recent work in assessment research.

Historical

The first known examinations were those for entry into the Chinese imperial civil service that date to 2000 BCE. These exams were an essential requirement for those wishing to achieve political or economic power. The examinations lasted two days and were a direct means of political control by the emperor, a Machiavellian system of ‘adult learning’ that remained in place until the mid-twentieth century, which makes Scotland’s examination system young by comparison, despite its status in western educational history.

As the general histories of Thomas Devine, and the specialist work of Walter Humes and T K G Bryce demonstrate, mass examinations were (along with mass education) a 19th century phenomenon that, as well as rewarding and elevating the able, were designed to separate pupils out into trades, professions and vocations. ‘The main factor in the future prosperity of nations was ‘the disciplined intelligence of the great bulk of the community’.’ (Devine, p 401) But many examinations had a mild eugenic function, sifting pupils destined for manual labour and trades from those with a more ‘academic’ bent.

As O’Connor and Robertson outline in their article *The Setting Up of the Scottish Leaving Certificate* (1990), 1867 was the watershed year for Scottish examinations, with the Royal Commission of Education in Scotland undertaking the first major review of state education in its then 200 year history. This led to the Education Act of 1872, its object to provide education for ‘the whole people of Scotland’. The Act saw the introduction of the Leaving Certificate Examination, a common examination for the many schools (including the ‘Higher Grade’ schools) that made up the Scottish system. Certificates were awarded on the result of a purely external examination for passes in the separate subjects of instruction. These would be rationalised into a single certificate that marked the successful close of the Intermediate and post-Intermediate stage respectively.

But it is in recent history that Scottish Education has undergone its most significant changes — from Comprehensives to Higher Still and the incorporation of FE colleges. As the various studies produced by the CES, Scottish Executive, HSDU and SFEU on Higher Still show, this transformation has taken place over a relatively short period of time, with teachers and lecturers having to keep pace with new demands on assessment and changing frameworks.

Higher Still, and the long-term move towards unitisation, had major implications for assessment in Scotland. The annual surveys of Higher Still's implementation, carried out by SFEU and the CES, were highly useful longitudinal measures of the effects of new assessment approaches, burdens and the effects of unitisation — a rare opportunity to observe a system as it emerges. A joint paper with SQA (Hart and Howieson) discusses the many challenges unitisation posed for the qualifications and examinations system during its development, as well as the implications for its future.

Titles

- ◆ *School Examinations, Encyclopaedia Britannica, 2005*

A straight history of examinations as a social phenomenon from early times to the present; not for the specialist, but provides a useful (and perhaps sobering?) overview of the many different methods, approaches and attempts to develop a standardised measure of academic ability and achievement in human history. The encyclopaedia also has detailed essays on the history of vocational and occupational learning, from medieval guilds and the production of 'masterpieces' to the apprenticeships of the industrial era.

- ◆ Devine, T M. (1999) Educating the People (Chapter 17) *The Scottish Nation 1700–2000*. Penguin Books.

A useful background for the development of secondary education, but assessment is largely treated in passing.

- ◆ Paterson, Lindsay. (2003) *Scottish Education in the 20th Century*. Edinburgh: Edinburgh University Press.

An important survey by Professor Paterson that has as its main aim an attempt to understand the alleged 'decline' of Scottish education. Assessment and examinations are naturally included in his deliberations.

- ◆ Humes and Bryce. (2003) *Scottish Education, Post-Devolution* (2nd ed). Edinburgh: Edinburgh University Press.

This book is essential reading for those wishing to grasp the systemic, philosophical, practical and political features that make Scottish education unique.

Of particular relevance to assessment are:

- ◆ Bryce, T G K and Humes, W M. (2003) *An Introduction to Scottish Education: Post-Devolution* (2nd ed) (Chapter 1). Edinburgh: Edinburgh University Press.

A useful, concise survey of Scottish education's main features, including its methods of assessment.

- ◆ Bryce, T G K and Humes, W M. (2003) *Scottish Secondary Education: Philosophy and Practice (Chapter 4) Scottish Education: Post-Devolution (2nd ed)*. Edinburgh: Edinburgh University Press.

Assessment theory, philosophy and practice are covered in this general essay on the ideas and ideals underlying the modern Scottish system.

- ◆ Humes, W M and Bryce, T G K. (2003) *The Distinctiveness of Scottish Education (Chapter 10) Scottish Education: Post-Devolution (2nd ed)*. Edinburgh: Edinburgh University Press.

The Scottish examination framework is one of the many features which make the overall system distinctive, a fact discussed in some detail by the authors. The traditional broad base of the curriculum is reflected in the logistics of examinations, moderation and assessment methods.

- ◆ Bryce, T G K. (2003) *Could Do Better? Assessment in Scottish Schools (Chapter 78) Scottish Education: Post-Devolution (2nd ed)*. Edinburgh: Edinburgh University Press.

In part, a history of the assessment process in Scotland, Bryce's essay is also an analysis of Scotland's current examination system, and is a useful primer for those attempting to understand the political processes and tensions behind contemporary educational policy and development. The piece tends towards the polemic, arguing for significant changes to assessment practice in Scotland.

- ◆ T G K Bryce and W M Humes (eds), Raffe, D. (2003) *Findings on Participation and Attainment in Scottish Education (Chapter 89) Scottish Education: Post-Devolution (2nd ed)*. Edinburgh: Edinburgh University Press.

The latest CES (Centre for Educational Sociology) research, incorporated into the new post-devolution edition of *Scottish Education*.

- ◆ SFEU/HSDU. (2000) *Implementation Studies in Colleges — Higher Still in Practice: Survey Report (Year 1)*. Scottish Further Education Unit.

The first of a series of reports that analysed the implementation of Higher Still in colleges. Each of the colleges was asked to fill in an annual survey, covering all aspects of the framework, from curriculum design to assessment. Their responses were crucial in informing the progress of Higher Still in colleges and the development of assessments for the new system. Feedback on assessment for Project-based National Courses proved particularly important.

- ◆ SFEU/HSDU. (2001) *Implementation Studies in Colleges — Higher Still in Practice 2000-2001*. Scottish Further Education Unit.

The second report on Higher Still implementation was able to build upon the previous survey, ask more targeted questions, and update on outstanding issues from the previous year.

- ◆ SFEU. (2002) *Higher Still in Practice — Survey Report (Year 3)*. Scottish Further Education Unit.

The last SFEU-published report on the annual Higher Still Survey.

- ◆ Raffe, D, Howieson, C and Tinklin, T. (2004) *The Introduction of a Unified System of Post-Compulsory Education in Scotland*. End of Award Report to the ESRC. Edinburgh: Centre for Educational Sociology, University of Edinburgh.

Arguably one of the most significant reports in Scottish Educational Research, this report is a retrospective on the development of Higher Still, evaluating its success and failures. The issue of assessment was a recurring debate throughout the process, and these discussions are summarised in this report.

- ◆ Hodgson, A, Howieson, C, Raffe, D, Spours, K and Tinklin, T. (2004) Post-16 Curriculum and Qualifications Reform in England and Scotland: Lessons from Home International Comparisons. *Journal of Education and Work*. 17(4).

The CES study provides a useful comparison of Scottish and English educational reform.

- ◆ Hart, J and Howieson, C. (2004) Unitisation — Benefits and Issues, *Research and Information Services Bulletin No 9*. Glasgow: SQA.

A unitised system holds many implications for assessment — particularly with the capacity for individual assessment and grading of each Unit. This paper is a useful historical study and highlights many important assessment issues for the future of the system in Scotland.

- ◆ Scottish Executive. *Assessment, Testing and Reporting: 3–14 Consultation on Partnership Commitments*. Consultation Paper. Scottish Executive.

This consultation formed a crucial part of the wide-ranging and far-reaching 3–14 review. The focus of the report is mainly structural and infrastructural, looking at the constituent partnerships, understandings and arrangements that sustain the current system.

- ◆ Scottish Executive. (2003) *Review of Modern Apprenticeships*. Scottish Executive: Central Research Unit.

Essentially an evaluative document that would win the case for further support and continuance of a successful scheme, the review document also provides a useful potted history of the development of this scheme. Assessment methodology — particularly work-based — forms an important part of the commentary.

2 Assessment theory and practice — general

Commentary

The theory of assessment has historically been less well developed than other forms of educational theory. As a result, it is more commonplace to speak of assessment principles rather than theories. Such principles developed out of actual practice, the accumulated knowledge and experience of generations of teachers and lecturers, typified in Brown et al's *500 Tips for Assessment*. Yet there is some evidence of a more concerted theoretical approach in recent years. A catalyst for this has been new technologies and means of administering assessment, and much of the best available literature on assessment principles has been produced by CAA researchers who, for obvious reasons, have been required to review assumptions about assessment practice in combining it with information technology.

Mhairi McAlpine's 2002 paper *Principles of Assessment* (CAA Centre, 2002) is a useful introduction to the overall subject, laying out the essential terminologies and classifications and discussing, in some detail, issues of validity. For McAlpine, assessment must be understood, first of all, as a form of communication, primarily between student and teacher but also to employers, curriculum designers and policymakers. Assessment is thus a social function, a communications link between the education system and wider society. Taking communication as her model, McAlpine then breaks assessment down into five main points or criteria that ensure that this dialogue is worthwhile: clarity of purpose; validity/reliability; referencing (establishing a common measure across all candidates); the quality of assessment items/instruments; and grading, a process that relates directly to referencing.

McAlpine also points out that assessment styles and approaches are subject to national variation. SQA uses criterion-referenced assessment, assessing candidates according to certain predetermined performance indicators and based on evidence they produce, and presumes that a student will pass examination. Failure to pass is seen as a fault of the learning process, not the individual.

Until the 1970s, the rest of the UK and the US predominantly opted for norm- or cohort-based assessment that assesses children relative to their peers or a predefined group. Candidates are ranked and compared with one another, the evidence showing how well they can do a task or understand a subject relative to everyone else.

McAlpine also explains formative and summative assessment (see below), but also defines convergent/divergent assessment. Convergent and divergent assessment are not so much separate types of assessment as different applications within formative and summative approaches, performing opposite

functions. Convergent assessment takes a predetermined ‘thing’ and sets out to discover whether a learner knows, understands or is able to do it. Divergent assessment is an open-ended process that aims to find out what the learner can do. While on the surface convergent assessment seems naturally suited to summative purposes and divergent to formative, in truth, one can find aspects of them in either.

Of all the jargon and terminologies used in assessment literature, validity and reliability are of paramount importance. Validity seems an especially difficult concept to pin down to specifics, and some literature seems a little cavalier in its use. SQA uses the traditional definition of validity in designing its assessments, putting in place a number of checks to ensure it continues to be ‘an inherent component’. Alternative interpretations suggest that assessments do not possess inherent validity but that it is defined by the use to which their results are put (sometimes termed consequential validity).

A reliable assessment is one that is judged consistent across all assessors and all candidates undertaking the same assessment task. Reliability depends on the validity of the assessment, but also on ample evidence of a student’s achievements, and on clearly defined Performance Criteria.

There are many debates over how assessment should be designed and administered, and this leads invariably to issues of curriculum design and policy. Assessment burden is one of these — how much is too much; to what extent does it deter learning? Does it motivate it? Literature from America, including an accessible discursive piece by Alexander Russo of the periodical *Catalyst*, often describes a ‘test glut’ that overburdens teachers and does not even give them useable data to improve school programmes. In a SAUM (Supporting Assessment in Undergraduate Mathematics) paper in 2001, Madison gives a useful summary of how the word has in itself functioned as a barrier, and in a few sentences, gives a concise summary of the many attempts to clarify the process:

‘We distinguished between summative assessment and formative assessment to try to clarify why assessment is done. We resorted to assessment cycles to imply that assessment was a continuous process rather than a discrete event. We added prepositional phrases to clarify the purpose when we talked of assessment of student learning and assessment in the service of learning. We tried to distinguish kinds of assessment by referring to classroom assessment, large-scale assessment, authentic assessment, and alternative assessment ...’

Yet, despite contributing greatly to the size of its dictionary entry, a definitive view of assessment is elusive; the very word continues to excite debate, even foreboding, amongst teachers and students alike. Debates over Higher Still have focused on the level, types, and quantity of assessment. The actions of the National Qualifications Task Group (NQTG) in opting to reduce the level of assessment in NQ suggests a developing consensus that assessment can be a barrier if used to excess.

The *14–19 Education and Skills* White Paper has also recommended reducing the number of A-level assessments from six to four. There is no question that much of the literature seems to *assume* a burden; yet there is a definite gap for research into how this ‘assessment burden’ actually affects learners.

In his chapter for the book *Agenda for a New Scotland* (January 2005), Lindsay Paterson sums up some of the blame that is apportioned towards assessment by critics of the educational system:

‘One is the sense that, in creating opportunity for all, we have ended up imposing impossible amounts of external assessment on all. The objection here is not to the intellectual challenge of examinations, but rather to the educationally harmful effects of too intrusive assessment, and also to the trivialising effects of assessment that is about competence and performance rather than understanding and knowledge.’

The most contentious of Paterson’s arguments is probably this last, over the place of competence in assessment criteria. The term ‘trivialising’ is highly pejorative, and many who have been active in educational policy and development may beg to differ.

Titles

- ◆ (1994) Assessment, *Encyclopaedia of Education*. MacMillan.

Classifies all of the major concepts and principles of assessment theory and practice. A useful, if generic, primer.

- ◆ Scottish Executive. (2005) *Circular No 02, June 2005: Assessment and Reporting 3–14*. Scottish Executive: Education Department.

An uninspiring title, but this is a significant document that lays out the Executive’s post-3–14 intentions for assessment in the Scottish curriculum.

- ◆ Scottish Executive. (2004) *A Curriculum for Excellence*. Scottish Executive: The Curriculum Review Group.

The report of the curriculum review, laying out their proposals for the curriculum, including assessment.

- ◆ McAlpine, Mhairi. (2002) *The Principles of Assessment*. University of Luton: CAA Centre.

Although produced by a CAA research unit, this useful paper is a concise, unpretentious primer in current assessment theory and practice. It outlines all of the major types of assessment and gives a straightforward definition of its essential principles.

- ◆ Scottish Executive. (2001) *Report of the First Survey of Modern Languages (French and German)*. Scottish Executive: Assessment of Achievement Programme.

Modern Language programmes often pose considerable assessment challenges, and this report looks at the various issues affecting assessment of French and German. While obviously subject specific, there is much here that is transferable to other disciplines and curricula.

- ◆ T G K Bryce and W M Humes (eds), Croxford, L. (2003) *Baseline Assessment in Scotland* (Chapter 84) *Scottish Education: Post-Devolution* (2nd ed). Edinburgh: Edinburgh University Press.

This chapter discusses the practice and development of baseline assessment in Scotland, outlining the debates, frameworks, theoretical and policy issues in turn. A useful introduction.

- ◆ *Assessment 5–14: What do pupils and parents think?* SCRE, Spotlight 87.
- ◆ Scottish Executive. (2002) *NQTG, Consultation on review of assessment within New National Qualifications: Summary of Analysis of Responses*. Scottish Executive

The consultation found that few respondents supported drastic changes to the status quo or to Higher Still, but supported a balance of internal and external assessment with a reduced volume of assessments.

- ◆ Scottish Executive. (2004) *Research on Transition from Primary to Secondary School*. Scottish Executive.

The focus and scope of this research is to be clarified in the Scottish Executive's new research programme but might involve using ScotXed data to analyse pupil attainment at the transfer from P7 to S1.

- ◆ HMSO. (2004) *The Tomlinson Report*. London: HMSO.

Proposals for the biggest shake-up of the qualifications and examination system in England make this required reading for an understanding of current trends and attitudes to assessment policy. The plans from the Working Group on 14–19 Reform — led by Mike Tomlinson, former chief inspector of schools — are expected to recommend a programme to transform the current system of GCSEs, A-levels and vocational qualifications into a new diploma qualification over 10 years.

- ◆ HMSO. (2004) *14–19 Education and Skills*. White Paper. London: HMSO. Among many of the proposals in this paper are changes to assessment regimes, including reduced assessment in A-level programmes.

- ◆ Brown, Sally. (1996) *Assessment*. Discussion paper. University of Northumbria.

Dated, but a useful measure of current assessment practice in HE.

- ◆ Brown, S, Race, P, and Smith, B. (1996) *500 Tips on Assessment*. Kogan Page.

A book that attempts to draw together the fundamental values underpinning assessment nine years on, and still provocative and informative.

- ◆ Paterson, Lindsay. (January 2005) Education: The Importance of Culture, *Agenda for a New Scotland*. Edited by Kenny MacAskill MSP.

A critical essay that discusses the effects of assessment on the Scottish education system, with some controversial conclusions.

- ◆ Russo, Alexander. (February 2004) *Test glut a burden to preschools*. Chicago USA: Catalyst.

An interesting article in that it discusses the on-going debate in the US over the 'test glut'. American education is known for the importance it places on standardised and regular classroom testing in particular, and as this article confirms, the debate continues to attract controversy.

- ◆ Madison, Bernard L. (2001) *Assessment: The Burden of a Name*. SAUM paper. University of Arkansas.

A discussion of assessment's 'image problem', which often belies its true nature, the paper suggests eradicating a separate concept of 'assessment' altogether. As assessment is, supposedly, an integral part of teaching and learning, there is no need to distinguish it. A weaker suggestion of his is to rename it — a much tried tactic that has a limited, or non-existent, shelf life.

- ◆ Race, Phil. (1995) The Art of Assessing, *The New Academic*. Vol 5, issue 3. SEDA.

A lively discussion of assessment techniques in higher education.

- ◆ Raffe, D. (2003) Examining the System, *Holyrood Magazine*. Issue 94, *SQA Supplement* (p13) (8 September 2003).

- ◆ Brown, G I, Bull, J, and Pendlebury, M. (1997) *Assessing Student Learning in Higher Education*. London: Routledge.

Discusses the theory of assessment as applied to an HE or HEI environment.

- ◆ (2002) *Report of the First AAP Survey of Social Subjects Enquiry Skills*. Assessment of Achievement Programme.

- ◆ Psychometric Tests, Liam Healey and Associates Assessment and Development Centres:
<http://www.psychometrics.co.uk/adc.htm>.

A useful background to the principles behind psychometric and personality tests.

- ◆ Black, Paul and Wiliam, Dylan. (October 1998) *Inside the Black Box: Raising Standards Through Classroom Assessment*.
<http://www.pdkintl.org/kappan/kbla9810.htm>
 Prentice-Hall Companion Website for Student-Involved Assessment for Learning, 4th edition (http://wps.prenhall.com/chet_stiggins_student_4).
 Argues for the potential of classroom-based assessment and its impact on raising standards and quality in teaching and learning.

- ◆ Sparks, Dennis. (Spring 1999) Assessment without Victims: An Interview with Rick Stiggins, *Journal of Staff Development*. 20 (2). Retrieved February 5 2005 from:
<http://www.nsd.org/library/publications/jsd/stiggins202.cfm>.
 Interview with the leading American assessment researcher.

- ◆ Johnson, M. (1977) *Intentionality in Education*. Albany NY: Center for Curriculum Research and Services, State University of New York.
<http://acase.org/mjohnson/>.
 Johnson has a loyal following among American educators and researchers, many of them having attempted to realise his theories of curriculum structure in practice. *Intentionality* is thus essential reading to understand some of the more radical ideas in education, but is not among the more accessible of educational texts. Assessment is seen as part of a cooperative approach to educational development. ‘Evaluation’ provides the essential raw materials for continuous development of the curriculum.

- ◆ Alloway, T P. (2003–2004 — on-going) *Development of a School-Based Tool for Assessing Working Memory*. Teaching and Learning Research Programme, ESRC.
 Working memory refers to the capacity to store and manipulate material over short periods of time, and has obvious implications for both performance in assessment and its application. A string of publications can be expected on this subject over the next period by researchers largely based in the University of Durham.

3 Formative and diagnostic assessment

Commentary

Teaching and learning are reciprocal processes that depend on and affect one another. Thus, the assessment component deals with how well the students are learning and how well the teacher is teaching.

(Kellough and Kellough, 1999)

Formative assessment has attracted increasing interest as a topic of research, classroom practice and educational policy. The potential of assessment as a means of guiding the teaching and learning process has excited many, although is not, of course, entirely new. The Qualifier Exams, Eleven Plus, and the entrance examinations for universities and private schools arguably had some formative element, though such tests were often a surreptitious means of selection and elimination.

As the pioneering work of Black and Dylan argued, genuinely formative assessment must be used to guide the learning process, a means of acquiring information and observations about a learner's skills and abilities to target practice better. In an era when educational resources are tight and must be deployed effectively while paying due attention to quality, formative theory is, understandably, an enticement to hard-pressed educators. But for learner and teacher alike, it seems a positive step to reconfigure assessment as more than just a means of acknowledging and measuring achievement, but also as an assist to higher levels of attainment. Diagnostic assessment has found its way into the curriculum, an initial stage in curriculum and programme design, a 'low' or 'no-stakes' test that does not count to grading, but provides a great deal of data for educators. Diagnostic tests are also being used and considered in higher education to assess whether a new student is able to cope with the demands of university study.

The work of Peter Honey and Alan Mumford is often cited in relation to diagnostic assessment, and although learning styles have come under some criticism, they remain a popular facet of diagnostic assessment. Thus, identifying a learning style becomes the primary purpose of the 'diagnosis'. But is this the same as identifying a viable curriculum?

The surest sign that an idea or practice has come of age is when policymakers devote significant research and development resources towards it. Assessment is for Learning was a major research project with a fairly self-explanatory title. The recent Executive response to the results of the consultation that followed the research confirmed their commitment to integrating formative assessment into the mainstream system — 'a streamlined and coherent system of assessment'. As they state:

‘The overall purpose of assessment to support learning is to provide all partners with sufficiently dependable information and feedback to inform judgements, choices and decisions about learning, and to inform planning for improvement.’

Titles

- ◆ Scottish Executive. *Personal Learning Programme 2002–2004 Evaluation Report*. Assessment is for Learning Development Programme: Scottish Executive.

This Scottish Executive publication focuses on the use of formative assessment as a developmental tool in the classroom, for directing learning and centring resources on the learner.

- ◆ Evaluation of Assessment is for Learning programme (Scottish Executive). The project which has been contracted involves the evaluation of the processes, products, and short-to-medium-term impacts of the Assessment is for Learning development programme. It is due to be completed in December 2004.

- ◆ Evaluation of Project 1 of the Assessment is for Learning development programme: Support for Professional Practice in Formative Assessment — Final Report, October 2004.

This report describes the evaluation of the first part of the Assessment is for Learning development programme: Support for Professional Practice in Formative Assessment, part of the Future Learning and Teaching (FLaT) programme. It investigated effective classroom approaches to formative assessment and produced guidance for schools and teachers about assessment policies and practice to raise attainment.

- ◆ Scottish Executive. (2005) *Assessment, Testing and Reporting 3–14: Our Response*. Scottish Executive.

This is a response by the Executive to the consultation that followed Assessment is for Learning, and is a useful indicator of their intentions towards assessment policy, in particular mainstreaming formative approaches into the curriculum.

- ◆ Black, Paul and Wiliam, Dylan. (1998) *Inside the Black Box*. Assessment Reform Group, King’s College University of London.

A pamphlet that gives practical advice and ideas for improving classroom assessment. The researchers were members of the Assessment Reform Group (ARG), an association of researchers dedicated to improving the general quality of assessment in the UK.

- ◆ Black, Paul and Wiliam, Dylan. (1999) *Assessment for Learning: Beyond the Black Box*. Nuffield Foundation.

A further ARG pamphlet that set out to reinforce to policymakers the messages implicit in the earlier research by Black and Dylan.

- ◆ Assessment Reform Group. (2000) *Assessment for Learning — 10 Principles*. Nuffield Foundation.

A short document outlining 10 principles for effective formative assessment.

- ◆ Honey, Peter, and Mumford, Alan. (2000) *The Manual of Learning Opportunities*. Peter Honey Publications.

Learning styles have become increasingly popular as a tool for teaching and learning. Honey and Mumford's questionnaire sets out to diagnose learning styles for each individual to better suit teaching methods to their personalities, aptitudes and inclinations.

- ◆ Swearingen, Richard. (2003) *A Primer: Diagnostic, Formative and Summative Assessment*. Washington USA: Heritage College.

A relatively jargon-free paper on diagnostic assessment and its uses in the classroom. Largely from an American context, but makes good use of international, especially Australian, research.

- ◆ Kellough, R D and Kellough, N G. (1999) *Secondary School Teaching: A Guide To Methods And Resources: Planning For Competence*. New Jersey: Copyright by Prentice Hill, Upper Saddle River.

Kellough and Kellough are frequently cited in American assessment literature, and their work has done a great deal to promote the concept of formative and diagnostic assessment, mainly through mainstreaming these ideas into textbooks such as this.

- ◆ Huba, M & Freed, J. (2000) *Learner-Centered Assessments on College Campuses: Shifting the Focus from Teaching to Learning*. Boston, Allyn and Bacon.

- ◆ *Formative and Summative Confidence-Based Assessment*. Proc. 7th International Computer-Aided Assessment Conference, Loughborough, UK, pp 147–155 (July 2003).

A useful summary of the latest thinking in confidence-based assessment. Somewhat academic, but good for references and generally clear and free of jargon.

- ◆ Black, Paul, Harrison, Catherine, Lee, Clare, Marshall, Bethan, and Wiliam, Dylan. (22 April 2003) *The Nature and Value of Formative Assessment for Learning*. Paper presented at AERA.
<http://66.102.7.104/search?q=cache:Ph-MeVrcqJsJ:www.kcl.ac.uk/education/papers/AERA%2520Pres.pdf+England+study+improved+formative+assessment+raised+student+achievement+overall+but+that+it+helped+low+achievers+most.&hl=en>

- ◆ Black, Paul and Wiliam, Dylan, (October 1998). *Inside the Black Box: Raising Standards Through Classroom Assessment* [Electronic version]. *Phi Delta Kappan*, 80 (2). Retrieved February 5 2005 from: <http://www.pdkintl.org/kappan/kbla9810.htm>
Prentice-Hall Companion Website for *Student-Involved Assessment for Learning*, (4th edition). (http://wps.prenhall.com/chet_stiggins_student_4)

- ◆ Stiehl, Ruth, Lewchuck, Les. (2005) *Tools for Building an Outcomes-Based College Curriculum*. League for Innovation in the Community College Learning Abstract, Volume 8 No. 2.

A somewhat complicated paper detailing how colleges can collaborate to redesign their curriculum to be more focused on student learning outcomes, including a convergent, cross-faculty approach to assessing learning.

- ◆ Stiggins, R J. (2002) *Assessment Crisis: The Absence Of Assessment FOR Learning*, *Phi Delta Kappan*. Retrieved February 5 2005 from: <http://www.pdkintl.org/kappan/k0206sti.htm>.

- ◆ Stiggins, R J. (2001) *Student-Involved Classroom Assessment* (3rd ed). Upper Saddle River, NJ: Prentice-Hall, Inc.

The relationship between formative and summative assessment — in the classroom and beyond, Classroom Assessment and the National Science Education Standards, Center for Education.

- ◆ Professor Keith Chapman. *Objective Tests in Formative and Summative Computer-Assisted Assessment*. University of Aberdeen: Department of Geography.
- ◆ Croxford, L. (2003) *The Early Intervention Programme in Scotland: Impact on the Attainment and Attitudes of Primary 3 Pupils*. Edinburgh: Centre for Educational Sociology, University of Edinburgh.

4 Summative assessment

Commentary

Two definitions of summative assessment:

Assessed work which contributes to the final outcome of a student's degree, such as unseen examinations, essays, dissertations or presentations.

(The University of Warwick 2004)

Summative assessment is testing which often occurs at the end of a term or course, used primarily to provide information about how much the student has learned and how well the course was taught.

(Wojtczak 2002)

When people think of assessment, it is usually in its summative sense, and despite the greater weight attached to formative, integrative and other varieties of assessment by contemporary educators, summative assessment has lost none of its importance. Indeed, with advances in Computer Assisted Assessment, the extension, expansion and convergence (between vocational and academic) of the curriculum, getting the summative stage of the assessment process right is arguably more important than ever.

Recent thinking attempts to link summative more clearly to formative assessment. For thinkers such as Johnson, the two formed the same evaluative process. A paper published by the Center for Education, *The Relationship Between Formative and Summative Assessment*, further explores this relationship between the two, seeing each as stages in a single learning process.

Given the often definitive nature of summative assessment, validity and reliability are the main stuff of debate — as in Bond's article on norm- and criterion-referenced testing and Chatterji's research on assessment design confirm. Moderation of summative assessment — which is often, though not exclusively, an external assessment — is also a vital issue, although much of the literature on this tends to be focused on technical guidelines and specifications, and seems under researched.

Titles

- ◆ Bond, L A. (1996) Norm and Criterion-Referenced Testing, *Practical Assessment, Research and Evaluation*. Vol.5 No.2.
<http://ericae.net/pare/getvn.asp?v=5&n=2>.
- ◆ (2003) *Formative and Summative Confidence-Based Assessment*. Proc 7th International Computer-Aided Assessment Conference. UK, Loughborough. July 2003. Pp 147–155.
www.ucl.ac.uk/~ucgbarg/tea/caa03.doc.

- ◆ McAlpine, Mhairi. (2002) *A Summary of Methods of Item Analysis*. CAA Centre Bluepaper 2, University of Luton.
<http://www.caacentre.ac.uk/>.

Item analysis is a vital component of quality control, usually in summative assessment, that works through looking at the constituent parts of an assessment. McAlpine gives a guide to the essential process of breaking an assessment into its components, and then analysing each of these separately to judge the overall quality of an examination.

- ◆ QCA. *Developing Assessment*. Innovating with History website:
<http://www.qca.org.uk/history/innovating/assessment.htm>.
- ◆ Chatterji, M. (2003) *Designing and Using Tools for Educational Assessment*. Boston: Allyn and Bacon.
A methodological survey.
- ◆ Gronlund, N. (2003) *Assessment of Student Achievement*. Boston: Allyn and Bacon.
- ◆ Wojtczak, A. (December 2000) *Glossary of Medical Education Terms*.
<http://www.iime.org/glossary.htm>. Revised February 2002.

Assessment in medicine is particularly geared towards summative assessment and medical education literature provides useful definitions of summative terminologies and methods. Wojtczak's book is one of the more recent glossaries and many of its definitions of assessment are largely generic.

- ◆ University of Warwick. *Course Specifications: Glossary of Terms relating to Course Specifications*.
<http://www2.warwick.ac.uk/institute/info/quality/coursespecs/view/glossary/>
Last revised: Tue Aug 24 2004.
- ◆ Stiehl, Ruth and Lewchuk, Les. (2005) Tools for Building an Outcomes-Based College Curriculum, Abstract, *League for Innovation in the American Community College*. Volume 8, Number 2.
- ◆ Johnson, D W and Johnson, R T. (2002) *Meaningful Assessment: A Manageable and Cooperative Process*. Boston: Allyn and Bacon.
- ◆ McKeachie, W J. (1999) *Teaching Tips: Strategies, Research, and Theory for College and University Teachers*. Boston: Houghton Mifflin Company.
- ◆ Popham, W J. (2000) *Modern Educational Measurement: Practical Guidelines for Educational Leaders*. Boston: Allyn and Bacon.

- ◆ Trice, A (2000) *A Handbook of Classroom Assessment*. New York: Addison Wesley Longman Inc. Aiken, L R. (2000) *Psychological Testing and Assessment (10th Edition)*. Boston: Allyn and Bacon.
- ◆ Angelo, T A. (Summer 1991) *Ten Easy Pieces: Assessing Higher Learning in Four*.
- ◆ Page, Gordon, Pachev, George, & Schreiber, William. *The Validity of PBL Tutors' Summative Assessment of Students*. Faculty of Medicine, University of British Columbia

This paper looks at the relationship between formative and summative assessment — in the classroom and beyond, Classroom Assessment and the National Science Education Standards as set by the American Center for Education.

The paper also discusses the relationships between two seemingly opposite forms of assessment. It emphasises the importance of linking both approaches within a curriculum, and of the comparability of their underlying principles.

- ◆ The Standards Site:
<http://www.standards.dfes.gov.uk/primary/features/literacy/supportforyear2/1128137>.
- ◆ Chapman, Keith. *Objective Tests in Formative and Summative Computer-Assisted Assessment*. University of Aberdeen: Department of Geography.

5 Integrative assessment

Commentary

In the US, there are numerous schools of thought on assessment and assessment practice — to be expected in a country that sets such importance on SATs. Particularly influential is the work of Mauritz Johnson, whose *Intentionality in Education* proposed an ‘evaluation schema’ for a continuous and integrated model of assessment and pre-assessment that is marbled throughout the learning process. This was taken up by adherents such as Paul Zachos who went on to develop a complete model of evaluation processes for developing curricula. In integrative assessment, local level research and evaluation is crucial to its development as an intrinsic part of the learning provision.

Also interesting is Gail Jansen van Rensburg’s case study of integrating assessment into a computer competency module. As integrative assessment works at a ‘micro-level’ (the classroom) it is often through case studies that the various issues are highlighted.

Integrative assessment has many affinities with formative assessments, due to its role in planning and developing teaching programmes.

Titles

- ◆ van Rensburg, Gail Jansen. *Developing Competent Learners through Integrative Assessment*. South Africa: Vaal Triangle Teknikon.
This study used action research cycles over six semesters to develop an entirely integrated method of formative, summative and pre-assessment for the curriculum.
- ◆ Johnson, Mauritz. (1977) *Intentionality in Education*. Albany, NY: Center for Curriculum Research and Services, State University of New York.
<http://acase.org/mjohnson/>.
- ◆ Zachos, Paul. (2004) *A Model of Evaluation Processes in Education: Based on Johnson’s Evaluation Schema*. New York: The Association for Cooperative Advancement of Science and Education.
- ◆ Furger, Roberta. (2002) Assessment for Understanding, *Edutopia Magazine*. 21 January 2002. http://edutopia.org/php/article.php?id=Art_937&key=005.

6 Teacher assessment

Commentary

Teachers often issue their own assessments as part of the on-going learning process, and there is a growing literature on how best to integrate and formalise these processes in the classroom. Teachers assess pupils on ability, behaviour and possibly learning difficulties and use the information gained to focus their efforts. As Claire Elwell's article *Teacher Assessment in the Secondary School* in the journal *Dyslexia Review* explains, teacher assessment can also be a catalyst for bringing in outside support and opinions, which make teacher assessment sound very like formative. Teacher assessment, however, tends towards the summative, and was defined by the Scottish Executive as an internal form of summative assessment (Circular 02, *Assessment and Reporting 23–14*, June 2005). This only confirms that the apparently opposite poles of formative and summative are hardly as distant as one might be led to believe.

Teacher assessment seeks to build up a profile of each pupil based on strengths and weaknesses and often covering a number of subjects. Teacher assessment is also incorporated into longitudinal approaches to student performances; the UK DfES has published several papers and online guides to aid teachers in tracking pupils' progress.

Of course, teacher assessment also occurs in other educational sectors. In the college sector FENTO (Further Education National Training Organisation) standards are an important aspect of FE lecturer training programmes. Tummons's practical text *Assessing Learning in Further Education* was written in response to the development of FENTO national standards for teachers and trainers in further education, assessment theory and being integral to FENTO endorsed qualifications. It is a useful text for its explanation of how theory relates to practices, and works as a practical study aid for lecturers.

Most of the literature on higher education tends to comment on teacher assessment in schools, but a cluster of papers (mostly for the Staff and Educational Development Association) from the early 90s by Race and Brown were important in starting academic debate on assessment as a whole, and opening up the subject of assessment in higher education for further research. Working out of the University of Northumbria, and in a culture that prioritised teaching over assessment, Race and Brown were successful in bringing the issue to the fore and shifting the focus of educationalists away from theory that could be applied to others, towards techniques that they themselves, as teachers in HEIs, could learn from. In general though, teacher assessment in the context of higher education seems under researched — something that could be said about assessment in higher education in Scotland overall, the issue being skirted in the recent *National Education Dossier*.

Titles

- ◆ Scottish Executive. (2005) *Assessment and Reporting 3–14*. Circular No 02 June 2005. Scottish Executive Education Department.

Uninspiringly titled, but this is a significant document that lays out the Executive's post-3–14 intentions for assessment in the Scottish curriculum.

- ◆ Tummons, Jonathan. (2005) *Assessing Learning in Further Education — Meeting the National Occupational Standards*. Learning Matters, York College.

Tummons's new guide to teacher assessment in FE looks at how (and whether) FENTO standards relate to the everyday experiences of FE teachers. It also summarises relevant research and theories and includes practical exercises in assessment practice for trainees.

- ◆ Brown, Sally. (1996) *Assessment*. Discussion paper, University of Northumbria.

Brown was the Director of Quality at the University of Northumbria in 1996, and this article in many respects sets forth her intentions for the role. The ideas and views she expresses are deliberately provocative; the intention to stimulate debate on current assessment practice in HE and how it should change. Dated, but an interesting yardstick against which the present day can be judged.

- ◆ Race, Phil. (1995) The Art of Assessing in *The New Academic*. Vol 5, issue 3. SEDA.

'The most important thing we do for students' says Race 'is assess their work.' A lively discussion of assessment techniques in higher education.

- ◆ Donaldson, A J M, Topping, K J. Resources for self and peer assessment, in *Promoting Peer Assisted Learning Amongst Students in Higher and Further Education*. Paper 96. SEDA.

A practically focused paper, providing criteria checklists for assessment in further and higher education

- ◆ Race, Phil. Quality of Assessment, from *Never Mind the Teaching Feel the Learning*. Paper 80. SEDA.

Race proposes ten 'worries' about assessment, and explores ways of dealing with them to improve the quality of assessment.

- ◆ Elwell, Claire. (1997) Teacher Assessment in the Secondary School in *Dyslexia Review*. Summer Issue.

A rather pedestrian case study of teacher assessment as applied in selective all-boys school; the main focus is on identifying special needs, although it makes several useful general points.

- ◆ DfES. (2005) *Using teacher assessment to track pupils' progress in mathematics*. Key Stage 3 National Strategy Site:
http://www.standards.dfes.gov.uk/keystage3/respub/ma_teach_assess.

Provides a set of materials to aid English and Welsh teachers to track pupils' progress in Maths.

- ◆ De Bota, Jennie. (2004) *Report on Electronic Teacher Assessments Project 2004*. Surrey County Council/Runnymede Centre.

A report on a project to exchange electronic teacher assessments for pupils at the end of both Key Stages 2 and 3 in the summer term of 2004 between institutions.

- ◆ Teasdale, A, Lueng, C. (2000) Teacher assessment and psychometric theory: a case of paradigm crossing?, *Language Testing*. Vol 1, Number 2, April 2000, pp 163–184. Hodder Arnold.

This article, based on research in England and Wales by academics from London's King's College, is an interesting critique of the use of psychometric tests in the classroom.

- ◆ Arkoudis, Sophie, O'Loughlin, Kieran. (2004) Tensions between validity and outcomes: teacher assessment of written work of recently arrived ESL immigrants, *Language Testing*. Vol 21, Number 3, 1 July 2004, pp 284–304. Hodder Arnold.

An Australian ESL study that investigates the struggle to maintain validity amid the tensions that develop naturally between teachers and state-led guidelines and initiatives.

- ◆ Hounsell & Entwistle. (2001–04) *Enhancing Teaching-Learning Environments in Undergraduate Courses*. Teaching and Learning Research Programme, ESRC (2001–04 — to be published).
- ◆ Brennan, Jary, Richardson & Osborne. (2004–07 — on-going) *What is Learned at University: The Social and Organisational Mediation of Learning*. Brennan, Jary, Richardson & Osborne Teaching and Learning Research Programme, ESRC.
- ◆ Fuller, Healey, Hurst, Riddell & Wareham. (2004–07 — on-going) *Disabled Students' Learning in Higher Education*. Teaching and Learning Research Programme, ESRC.

7 Computer Assisted Assessment

Commentary

Moves to utilise the capabilities of ICT in administering and creating assessments have had one of the greatest single impacts on assessment literature in recent years. The process requires educators and researchers to rethink — even relearn — aspects of their assessment practice — for no amount of technical innovation can replace solid assessment principles and theory. Mhairi McAlpine's CAA Bluepaper *The Principles of Assessment* reinforces this need to link innovation with tried and tested assessment techniques. McAlpine re-examines and explains the principles of assessment component by component, so that when they are put back together electronically, assessment is both valid and reliable. In her second Bluepaper, McAlpine looks at item analysis, giving a methodology for this important aspect of assessment quality control. Even more practical is McAlpine's third Bluepaper, *Design Requirements for a Databank*, a guide to the design principles of this important CAA tool.

One of the more influential books on CAA is Bull and McKenna's *Blueprint for Computer Assisted Assessment*. The title is self-explanatory; this is a book of practicalities, focused on the implications and potential for assessment practice that is firmly grounded with its target audience. It is also the best way to get up to speed on the subject, as the authors review all of the major books and papers published on CAA, and base their advice and conclusions on research findings taken from a major government funded project. Bull and McKenna also carried out an extensive research survey of CAA use in higher education over 2003. Combining qualitative and quantitative methods, the survey was of some importance, but due to Bull's untimely death in 2003, was never published online. The paper is nevertheless listed for those wishing to secure a hard copy.

As with ICT itself, CAA is a fast-moving area. Papers, such as Chapman's 1997 case study of CAA in Geography, look slightly dated in light of recent leaps in the field. Currently, both the Scottish and UK administrations are piloting and rolling out CAA programmes in schools and colleges. As with the SFEFC report of 2003 (*Computer Assisted Assessment: Implementing CAA in the FE sector*), research effort is likely to turn ever more to the infrastructural and policy aspects of CAA. (See also McAlpine and Sclater, below).

Yet it is also likely that CAA will continue to produce highly technical literature for some time to come, adding an extra layer of ICT theory and logistics to the existing jargon, as illustrated by the chapter and article titles below.

Titles

- ◆ Ware, Martyn, McAlpine, Mhairi. (2004) *Introducing Computer Assisted Assessment: Laying the Foundations for an Integrated Approach*. SQA.

This paper lays out a blueprint for integrating CAA into the Scottish system. The main hope for the plan laid out by the authors is to avoid the ‘experimental wastage’ that habitually dogs CAA pilots. The document is a significant statement of intent; to work collaboratively across sectors and pool the resources of the major stakeholders in Scotland to make CAA a success.

- ◆ Herd, G, Clark, G. (2003) *Computer Assisted Assessment: Implementing CAA in the FE Sector in Scotland*. Report to SFEFC ICT Innovation Fund

As the title suggests, this report signified the acceptance of CAA (in Scottish FE) and signified wide-ranging plans to implement it across the sector. The report provides a useful update on progress in CAA and further links it to suggested actions and initiatives to integrate CAA across the educational system.

- ◆ McAlpine, Mhairi. (2002) *The Principles of Assessment (2002)*. CAA Centre Blueprint 1, University of Luton.
<http://www.caacentre.ac.uk/>.

An essential starter paper in CAA, and assessment in general, this is a useful primer in assessment methodology, outlining the principles and standards that must be maintained through the transition to ICT-based assessment.

- ◆ McAlpine, Mhairi. (2002) *A Summary of Methods of Item Analysis (2002)*. CAA Centre Blueprint 2, University of Luton.
<http://www.caacentre.ac.uk/>.

Item analysis is an important aspect of quality control usually in summative assessment; by looking at the constituent parts the overall quality of an examination can be judged.

- ◆ McAlpine, Mhairi. (2002) *Design Requirements of a Databank*. CAA Centre Blueprint 2, University of Luton (2002).
<http://www.caacentre.ac.uk/>.

- ◆ McKenna, Colleen, Loewenberger, Pauline, & Bull, Joanna. (2003) *Results of the CAA National Survey*. CAA Centre Blueprint 2, University of Luton.
<http://www.caacentre.ac.uk/>.

- ◆ Bull, Joanna, McKenna, Colleen. (2003) *Blueprint for Computer Assisted Assessment*. Falmer: Routledge.
<http://www.caacentre.ac.uk/>.

A second edition of the seminal book based on a government funded project, this is a practical, yet research-based guide for higher and further education practitioners who wish to use computers to assess their students’ learning. The book is a wide-ranging consideration of existing literature on CAA, but also

functions as an introduction to the practice of CAA, and generally avoids theorising or arguments ‘for or against’.

- ◆ Scottish Executive. (2005) *Impact of ICT Initiatives on Pupil Attainment, Interchange 63 and Insight 2*. Scottish Executive.

Two reports on the first two phases of research on the impact of ICT Initiatives in Scottish schools. A third phase of research to assess further impact of ICT initiatives and infrastructure was completed in February 2005.

- ◆ Herd, G, Clark, G. (2003) *Joined up? Or just lucky? Implementing CAA in Scotland*. Proceedings of the 7th Int. CAA Conferences. England: Loughborough University.
- ◆ McAlpine and Sclater. (2001) *The Value of Strategic Partnerships in Informing Learning Technology Development ALT-C*. Edinburgh: Edinburgh University.
- ◆ Ridgeway, Jim, McCusker, Sean. Report 10: *Literature Review of e-assessment*. School of Education, University of Durham, Daniel Pead, School of Education, University of Nottingham.
A useful and accessible review of current literature on CAA.
- ◆ McKenna, Colleen, Bull, Joanna. Quality assurance of computer assisted assessment: Practical and strategic issues, *Quality Assurance in Education*. No 8.
- ◆ Chapman, Keith. (1997) *Changing Assessment Practices in First and Second Year Geography Modules using Computer-Assisted Assessment (CAA) to set Formative and Summative Objective Tests*. University of Aberdeen.

Mostly of historical interest, recent advances in CAA make this paper look somewhat dated.

- ◆ Alloway, T P. (2003–04 — on-going) *Development of a School-Based Tool for Assessing Working Memory*. Teaching and Learning Research Programme.

Working memory refers to the capacity to store and manipulate material over short periods of time. This project aims to develop a computerised tool for assessing short-term and working memory for use with children aged four to eleven years (Automated Working Memory Assessment, AWMA). This tool provides a practical and convenient way for teachers to screen their pupils for significant working memory problems. The project will also advance theoretical understanding of the role of working memory in learning.

- ◆ Martin Bush. *Alternative Marking Schemes for On-Line Multiple Choice Tests*. Belfast: 7th Annual Conference on the Teaching of Computing, (provided with permission).

- ◆ Segall, Daniel O and Morena, Kathleen E. Development of the Computerized Adaptive Testing Version of the Armed Services and Vocational Aptitude Battery, *Innovations in Computerized Assessment*. Edited by Fritz Drasgow and Julie B Olson-Buchanan. Lawrence Erlbaum Associates, Publishers.
- ◆ Bergstrom, Betty A, Lunz, Mary E. *CAT for Certification and Licensure*. Edited by Fritz Drasgow and Julie B Olson-Buchanan. Lawrence Erlbaum Associates, Publishers.
- ◆ Kingsbury, C Gage, Houser, Ronald L. *Developing Computerized Adaptive Tests for School Children*. Edited by Fritz Drasgow and Julie B Olson-Buchanan. Lawrence Erlbaum Associates, Publishers.
- ◆ Mills, Craig N. *Development and Introduction of a Computer Adaptive Graduate Record Examinations General Tests*. Edited by Fritz Drasgow and Julie B Olson-Buchanan. Lawrence Erlbaum Associates, Publishers.
- ◆ Vispoeal, Walter P. *Creating Computerized Adaptive Tests of Music Aptitude: Problems, Solutions and Future Directions*. Edited by Fritz Drasgow and Julie B Olson-Buchanan. Lawrence Erlbaum Associates, Publishers.

8 Work-based assessment

Commentary

Interest in work-based assessment has intensified with the shift to bring academic and vocational education closer together. Work-based assessment takes place in the workplace, is often continuous and has a basis in competence and ability to perform work-related tasks and functions to standard — essentially a specialised form of convergence-based assessment (see above). But as the report by Gallacher et al on Modern Apprenticeship completion rates pointed out, good assessment should go beyond competences to advising the learner as to next moves and possibilities/opportunities for further development, and guide the training process. Simply assessing competences can amount to ticking a box and taking development no further. As the following quote from the report suggests, Modern Apprenticeships (through the VQ system) over-emphasise assessment:

‘I think if we are looking at the commercial training providers, they themselves would say quite openly, that they don’t consider themselves to be training providers any more. They consider themselves to be vocational assessment units. I think that’s probably fair comment.’

As Norcini’s article points out, ‘knows’ is the lowest assessment criteria on the work-based assessment scale in some professions. It should also be noted that there is a slight difference between work-based assessment and workplace assessment — the latter is specifically assessment that takes place on the ‘shop floor’.

Work-based methods have of course been around for some time but were often informal or an integral part of apprenticeships and ‘on the job’ training. Academic interest is therefore comparatively recent, but with the influence of *A Smart, Successful Scotland* and the culmination of Higher Still, SCQF and various moves to unify the system, increased academic attention has been inevitable.

The creation of the Centre for Research in Lifelong Learning was one of the major steps forward in encouraging work-based learning/assessment research in Scotland. The latest CRLI project is the on-going pilot study ‘The Implementation of Work-Based Learning within higher education and the Workplace’. The research will include a documentary analysis of the policy context for the emergence of work-based learning in higher education in the UK, a pilot study of two programmes, and a major research study of a range of practices. Its focus is mainly on learning, but assessment, formative or otherwise, also plays an important role. Research on New Deal in Scotland recommended more formative work-based assessment, prior to a learner’s entry into programmes.

As another CRLI publication by Cloonan and Turner highlighted, there have been some concerns over work-based assessment and its effect on overall

quality. Some LEC funded programmes, for example, become ‘output-related’ — or to be more specific, depend upon targets for payment. As Cloonan notes, work by Hyland and Matlay recommended in 1998 that work-based learning be removed from colleges altogether.

This, of course, has not happened, and current proposals and plans for the curriculum and the mapping of National Occupational Standards means that work-based assessment, in various forms, is here to stay.

Titles

- ◆ SFEU. (1999) *The New Deal in Practice: A Survey of New Deal in Scottish Further Education Colleges 1998–99*. SFEU.

An important study by SFEU which looked closely at New Deal provision in Scotland and made recommendations regarding assessment approaches and practices — particularly the move towards a more formative approach.

- ◆ Scottish Executive. (2003) *Review of Modern Apprenticeships*. Scottish Executive Central Research Unit.

Assessment methodology — particularly work-based — forms an important part of the commentary, with many responses outlining suggestions for changing assessment methods and, as with New Deal, moving towards a more integrated, even formative, approach.

- ◆ Fenton, Alison. (2004) *Case Study 3 — Developing Assessors at the University of Glasgow*. University of Glasgow paper.

Assessors are vital to Modern Apprenticeships and as Fenton’s short factsheet notes, one of the main difficulties they face is effective time management.

- ◆ FutureSkills Scotland. *Annual Employers Survey 2004*. FutureSkills Scotland.

A useful landscape survey in which to understand work-based assessment and training in Scotland, with figures outlining intensity of training among the workforce and whether this is work-based, day-release or off-the-job.

- ◆ Gallacher, Jim, Reeve, Fiona & Clarke, Julian. (2005) *The Implementation of Work-based Learning within Higher Education and the Workplace: a pilot study*. Glasgow Caledonian University/Open University: Centre for Research in Lifelong Learning.

A major study, still under way, of work-based learning (WBL) in higher education and workplaces that aims to increase understanding of the different forms of work-based learning.

- ◆ Gallacher, J et al. (2004) *Modern Apprenticeships: Improving Completion Rates*. Edinburgh: Scottish Executive Social Research Department.

In analysing how completion rates could be improved, this research also considers methods of assessment, including the performance of work-based

assessment and learning outcomes. It was found that the system of assessment, while in theory simple, involved a fairly Byzantine network of different agencies and individuals involved in assessment, and recommended measures to simplify and clarify roles and functions within the system.

- ◆ Cloonan, Martin, and Turner, Eileen. (2005) *The Past, Present and Future of Further Education in Scotland: A research-based guide to the literature for practitioners, (2000)*. Glasgow Caledonian and Stirling Universities: Centre for Research in Lifelong Learning.

While due for update, this is a useful general commentary of writing and research on FE in Scotland, summarising the main debates and controversies, including work-based learning and assessment.

- ◆ Rainbird, H. (2004) *Improving Learning in the Workplace, Teaching and Learning Research Programme Research Briefing No 7*. March 2004. Economic and Social Research Council.

More of a general overview of the environment in which workplace learning takes place, but assessment is also considered in this easily digested briefing.

- ◆ Rainbird, H. (2001) *Context, Power and Perspective: The Challenges to Improving Attainment at Work*. TLRP/SKOPE workshop held at University College Northampton, 8–10 November (2001): <http://www.tlrp.org/acadpub/Rainbird2001.pdf>.

An interesting paper discussing attainment in workplace learning environments, with assessment methods considered as a determining factor.

- ◆ CRL. (2003) *Proceedings of Experiential, Community, Workbased: Researching Learning Outside the Academy Conference*. Glasgow and Stirling: Centre for Research in Lifelong Learning.

Proceedings from the 2003 event on work-based and other non-academic forms of learning.

- ◆ Osborne, M, Gallacher, J and Cloonan, M. (1999) *Researching Lifelong Learning: An Introductory Briefing*. Glasgow Caledonian and Stirling Universities: Centre for Research in Lifelong Learning,.

A useful ‘beginner’s guide’, although now dated. Cites references to assessment research.

- ◆ Gallacher, J, Reeve, F. (2000) *Implementing Work-based learning in Higher Education: exploring control and negotiation in the curriculum*. Paper presented at the Working Knowledge Conference, Sydney, Australia. 10–13 December 2000.

- ◆ Norcini, J. (2003) *Work Based Assessment in ABC of learning and teaching in medicine*. British Medical Journal.

An interesting guide to work-based assessment as practised in the medical profession. Many of the issues are generic, and the theoretical framework Norcini describes is of interest to anyone engaged in work-based approaches.

- ◆ DfES. (2005) *Getting on in Business, Getting on at Work*. Government White Paper, HMSO.

Aims to build on the 2003 strategy, with further efforts promised to integrate workplace training and raise the skills level of the UK workforce.

- ◆ Hale Consulting. (2004) *Apprenticeships, Work-based learning and progression to Higher Education*. Summary of: Progression of Advanced Apprentices into Higher Education. Produced for West London Partnership.

9 Test/exam performance

Commentary

Performance in examinations/testing is more than just an educational issue; it is the area of education most likely to attract political and media interest. There are a number of reports, domestic and international, looking at the issue of examination and testing that are a far cry from simple 'league tables'. Perhaps the one that is of most interest to commentators and policymakers is the wide-ranging PISA study that takes in most of the OECD countries and regions, a cross-section of 15-year-olds, and their various proficiencies in core subjects. Such studies do not look at assessment alone, but ask questions of entire curricula and teaching standards, though assessment is an important issue, particularly where there is the possibility of national tests against the international, standardised tests of the PISA study.

The next survey will take place in 2006 and will concentrate on the testing of Science, but with some elements of English and Mathematics also being tested. The SCRE Centre provides the National Project Management for Scotland's involvement. This will involve sampling pupils and schools, administering tests, and analysing the results to allow international comparisons of student achievement to be made.

Domestic research also pays close attention to exam performance, some of which has a wider remit than educational interests. The 2002 research project to measure social deprivation in Scotland established clear links between poor performance and eventual social exclusion and poverty, and prompted changes to performance indicators.

Titles

- ◆ *Trends in International Maths and Science Study (TIMSS) (Sc Ex)*.
An international comparative study run by the International Association for the Evaluation of Educational Achievement (IEA) has been commissioned. This research assesses performance in Maths and Science of a sample of P5 and S2 pupils. An international report is due to be published in 2005.
- ◆ Fieldwork to support Scotland's involvement in the Programme for International Student Assessment (PISA), (SEED) (2006).
The Programme for International Student Achievement (PISA) is a high profile international survey of student achievement at age 15. The survey allows for comparative analysis of the performance across national borders through the administration of a standard test, and as a result raised many interesting questions about differing assessment regimes and the results they routinely deliver.

- ◆ Page, Gordon, Pachev, George, Schreiber, William. *The Validity of PBL Tutors' Summative Assessment of Students*. University of British Columbia: Faculty of Medicine.

A somewhat technical Canadian study of assessment validity/student performance in the medical/dental faculty, making use of ANOVA statistical testing.

- ◆ Howieson, C and Iannelli, C. (2004) *Scottish School Leavers Survey Special Study: The Effects of Low Attainment on Young People's Outcomes*. Report to the Scottish Executive. Edinburgh: Centre for Educational Sociology, University of Edinburgh.

This is a crucial report as it sets out to shed more light on the lost 50% of school leavers who do not re-engage with educational provision after leaving school. A hard-figures report that gives useful macro-level information on the subject, touching on assessment as an instrumental part of disengaging some learners — but also, in some cases, the last known contact before their 'disappearance'.

- ◆ (2002) *Assessment of Achievement Programme*. Report of the First AAP Survey of Social Subjects Enquiry Skills.

10 Gender

Commentary

‘Is gender a determining factor in assessment?’ This question is now being asked by researchers and practitioners, particularly as the performance of boys and girls in schools continues to diverge. But where do the roots of this divergence lie — in teaching and learning, in the contemporary classroom environment/culture (which is arguably much more feminine than the more ‘traditional’ models), or in the instruments used to assess students? As the Nuffield Foundation paper by Martyn et al makes clear, a number of assumptions on gender and assessment have been established; girls perform better in language-based subjects, boys in Maths and Science. Even with the overall outperformance of boys by girls, there is still significant discrimination against girls in the labour market.

Much of the published research focuses on this mismatch in performance in Maths subjects. Carr and Jessup’s research delved into the psychological and behavioural background of the different performance of boys and girls. It was found that their initial methods of understanding and learning were the route to much of these differences. Crane’s research also looked into the domestic ferment of children, and found a link with the cognitive capabilities of the mother, a theme also explored in the long-term study of Northern Irish pupils by Daly and Shuttleworth. It found that gender and school effects are small, and girls seem to be closing the performance gap on boys in these samples. Family socio-economic status and ability are the main factors affecting attainment and examination entry. The main findings of Fennema and Hart’s study were that gender differences may be decreasing, but still exist in the learning of complex Mathematics, personal beliefs in Mathematics and career choices that involve the subject, a finding confirmed in Friedman’s meta-analysis of gender differences in Mathematics.

The definitive text on the subject of gender and Maths is perhaps Leder’s *Mathematics and Gender*. Although a number of gender differences were found, the author offers some important qualifications; in the US, males participate more than females in non-compulsory Mathematics courses; the genders performed differently, especially in standardised tests, but all of these between-gender differences were dwarfed by much larger within-gender differences.

The Centre for Educational Sociology has engaged with the issue of gender in several reports and papers since 2001. A report to the Scottish Executive, *Gender and Pupil Performance in Scottish Schools*, undertook a major survey of gender as it influenced performance in the classroom and in assessments, although what it makes most clear is the need for further work in this area. There is no question that gender and assessment is an issue that is under researched, particularly with relation to the assessment methodology and the instruments themselves.

Titles

- ◆ Scottish Executive. *Evaluation of Initiatives to Address Gender Equalities in Performance*. Scottish Executive (to be completed autumn 2005).

This research focuses on identifying good practice and the effectiveness of initiatives developed by Education Authorities and schools to address gender inequalities in attainment.

- ◆ Andrews, Martyn, Bradley, Steve, Stott, Dave & Taylor, Jim. (2004) *The Educational Gender Gap, catch up and labour market performance*. Nuffield Foundation.

A useful summary of the main issues and assumptions on this issue.

- ◆ Burnett, P C. (1996) Gender and Grade Differences in Elementary School Children's Descriptive and Evaluative Self-Statements and Self-Esteem, *School Psychology International*. 17.

In a sample of elementary school children, boys reported higher scores on descriptive and evaluative statements on Mathematics.

- ◆ Tinklin, T and Croxford, L. (2000) Scottish School Leavers Survey: High Attaining Female School Leavers, *Scottish School Leavers Survey Special Report III*. Edinburgh: SEED.
- ◆ Tinklin, T, Croxford, L, Ducklin, A and Frame, B. (2001) *Gender and Pupil Performance in Scottish Schools*. Report to the Scottish Executive Education Department. University of Edinburgh: Centre for Educational Sociology.
- ◆ Tinklin, T. (2003) Gender Differences and High Attainment, *British Educational Research Journal*, 29(3). pp.307–325.
- ◆ Croxford, L, Tinklin, T, Frame, B and Ducklin, A. (2003) Gender and Pupil Performance: Where do the Problems Lie?, *Scottish Educational Review*. 35(2), pp 135–147.
- ◆ Tinklin, T, Croxford, L, Ducklin, A and Frame, B. (2001) *Gender and Pupil Performance in Scottish Schools*. Report to the Scottish Executive Education Department. University of Edinburgh: Centre for Educational Sociology.
- ◆ Carr, M and Jessup, D L. Gender Differences in First-Grade Mathematics Strategy Use: Social and Metacognitive Influences, *Journal of Educational Psychology*. 89(2).

Fifty-eight first grade children solved addition and subtraction problems individually and in groups of three in October, January and May. The children's strategy use was assessed — finger counting, memory, etc. Gender differences were found: girls were more likely to count on their fingers or use counters (overt strategies), boys were more likely to use retrieval (from memory) to solve addition and subtraction problems.

- ◆ (1997) *Scholastic Appeasement Test*. USA: George Mason University.
<http://www.stats.org/record.jsp?type=news&ID=7>.

‘Was there sex bias in the SATs?’ is the lead and leading question of this article, summarising the debate over the apparent inbuilt bias — or 35 point gap — between girls and boys in the Maths section of the test. The relationship between actual tests scores and predictions — or over predictions — makes interesting reading.

- ◆ Crane, J. (1996) Effects of Home Environment, SES, and Maternal Test Scores on Mathematics Achievement, *The Journal of Educational Research*. 89(5) pp. 305–314.

Using data from the US National Longitudinal Survey of Youth (over 7,000 children in the 1986 and 1988 waves), the relationship between Mathematics achievement and a number of background factors was tested. It was found that home environment, parental socio-economic status and maternal cognitive test scores all had a significant effect on children’s test scores at ages five and nine. It was found that about 25% of the relationship was due to maternal cognitive ability, while 75% was due to home environmental factors.

- ◆ Daly, P and Shuttleworth, I. (1997) Determinants of Public Examination Entry and Attainment in Mathematics: Evidence on Gender and Gender-type of School from the 1980s and 1990s in Northern Ireland, *Evaluation and Research in Education*. 11(2).

Using three samples of Northern Irish pupils taken in the 80s and 90s, the authors look at the effects of gender and co-educational versus single sex schools, taking into account pupil and family background factors such as religion, socio-economic status and verbal reasoning.

- ◆ Fennema, E and Hart, L E. (1994) Gender and the JRME, *Journal for Research in Mathematics Education*. 25(6) pp.648–659.

A review of the research on the relationships between gender and achievement in Mathematics. Gender differences in Mathematics vary by school, socio-economic status, ethnicity and teacher. Teachers tend to structure their classrooms in ways that favour male learning. Specific interventions can achieve equity in Mathematics.

- ◆ Friedman, L. (1989) Mathematics and the Gender Gap: A Meta-Analysis of Recent Studies on Sex Differences in Mathematical Tasks, *Review of Educational Research*. 59(2).

A meta-analysis of studies appearing between 1974 and 1986 found that gender differences in Mathematics were small, and were decreasing over the years. The difference in favour of boys is higher in the upper years of secondary school and among high achievers.

- ◆ Leder, G C. (1992) Mathematics and Gender: Changing Perspectives, *Handbook of Research on Mathematics Teaching and Learning*. Edited by D A Grouws. New York: MacMillan.

A comprehensive review of research on Mathematics and gender. The author also offers a number of possible explanations for the existing differences, including biological variables, environmental variables (school, teacher, peer group, parents and society), learner-related variables and cognitive variables (such as intelligence, spatial abilities, confidence, fear of success, attributions and persistence). Most important of these would seem to be subtle but consistent differences in cultural and societal pressures, and expectations for females and males which are frequently internalised by individuals and may lead to differing beliefs and expectations in areas critical to Mathematics learning.

- ◆ Massey, A J. (1997) Variations in Class Size and Achievement in GCSE Mathematics by Pupils from LEA/GM and Independent Schools, *Research in Education*. 57.

A sample (N=1170) of pupils taking the Midland Examination Group's GCSE examinations in June 1994 was studied to determine the relationship between class size, gender, LEA/GM status and GCSE Mathematics examination scores. In the basic tier, girls' performance was, statistically speaking, significantly (but not strongly) higher than that of boys.

11 Race and ethnicity in assessment

Commentary

Issues of race and ethnicity in assessment go beyond the straightforward specifics of language barriers and proficiency. The success of educational methods is to a great extent dependent on cultural factors, and assessment is no different, although in some researches (such as the controversial and widely discredited ‘Bell Curve’ research) cultural issues and prejudices are more marked in the manner of investigation itself rather than any valid findings.

Given the murky history of race and ethnicity in science and sociology, it need hardly be said that caution must be exercised in looking at this area. In 1992, Secada carried out a much needed and well-balanced study in Mathematics. The chapter discusses the relationship between Mathematics achievement and background factors such as ethnicity, race, language spoken and social class. Among US students, White students do better on standardised achievement tests than Hispanic students, who do slightly better than African American students. The gap in attainment between African American and White students narrows over time on basic skills, but not on higher level skills.

A large part of the ethnic differences mentioned above can be accounted for by differences in socio-economic status, as well as differences in language proficiency which also seems to be somewhat related to Mathematics achievement in that language. Classroom and school factors, such as different teacher expectations for different social groups and tracking, which might disadvantage low SES and ethnic minority groups, are also discussed. Direct instruction has been found to be an effective way of teaching basic skills to disadvantaged groups, as has cognitively guided instruction.

For Scots educators and researchers, the area is under researched, even as immigration to Scotland by those of different race, creed and ethnicity increases. The most logical place to start reading is the Race Relations Act, now coming into force in education. Previous exemptions no longer apply, and it will be required of all institutions that they do not discriminate, knowingly or otherwise, against candidates because of their background.

The issue is gradually being addressed — the recently announced (July 2005) consultation document, *Adult ESOL Strategy for Scotland*, has prioritised the need to effectively evaluate the existing skills and abilities of immigrants and ‘recent arrivals’.

Titles

- ◆ Secada, W G. (1992) Race, Ethnicity, Social Class, Language, and Achievement in Mathematics, *Handbook of Research on Mathematics Teaching and Learning*. Edited by D A Grouws. New York: MacMillan. Pp 623–660.
- ◆ Race Relations (Amendment) Act (2000). HMSO.
- ◆ CRE. (2005) *Learning for All*. Commission for Racial Equality. Sets racial equality standards in seven areas, including assessment and curriculum.
- ◆ Scottish Executive (2003). Education Department Race Equality Action Plan, Progress Report 2003.
- ◆ Scottish Executive. (2005) *Adult ESOL Strategy for Scotland*. Consultation Document, Scottish Executive/HMSO.

The consultation document proposes ways in which the various agencies, interests and institutions involved in ESOL can work better in meeting the needs of ESOL learners. Assessment methods must surely form a part of this debate?

12 Disability

Commentary

Disability issues in assessment invariably centre around practicalities; ensuring that a learner's disabilities do not disadvantage them in sitting assessments and examinations. The Beattie Committee was established in April 1998 to examine the needs of young people (ie those between 16 and 24) who had problems such as learning disabilities, mental health problems and feelings of alienation from society. Its remit was:

‘To review the range of needs among young people who require additional support to make the transition to post-school education or employment; the assessment of needs; the quality and effectiveness of current provision in improving skills and employability; and to make recommendations to improve coherence, continuity and progression.’

The implications of Beattie have been far ranging, and schools and colleges have been required to implement its recommendations and ensure that disability does not hinder learning or attainment.

Titles

- ◆ Scottish Executive. (1999) *The Beattie Report*. Advisory Committee on Post-school Education and Training for Young People, Scottish Executive.
- ◆ Tinklin, Teresa, Riddell, Sheila and Wilson, Alastair. Disabled Students in Higher Education, *Disabled Students and Multiple Policy Innovations in Higher Education*. CES Paper No 32.
- ◆ Riddell, S, Tinklin, T and Wilson, A. (2003) Disabled Students in Higher Education: Legislation, Teaching, Learning and Assessment, in M Slowey and D Watson (eds) *Higher Education and the Lifecourse*. Section 1, OUP.
- ◆ Fuller, Healey, Hurst, Riddell & Wareham. (2004–07 — on-going) *Disabled Students' Learning in Higher Education*. Teaching and Learning Research Programme, ESRC.

13 Comments and recommendations

On first glance, assessment would appear to be a well researched area in which many different avenues have already been explored. However, a simple comparison of Google search terms on teaching and learning versus assessment is a crude but effective means of illustrating the disparity — 76,100,000 results for teaching and learning research versus 40,100,000 for assessment research.

Numbers aside, the real issue for educational research in any field is surely the degree to which it is relevant, targeted and effectively related to practice. In this regard, organisations such as SQA are well placed to take a lead in ensuring existing research that meets these criteria is engaged with, and that research gaps are, where possible, identified and addressed.

General research on assessment theory and practice seems well established, comprehensive and robust, with a generally sound understanding of principles and theories. Research into computer assisted, summative and, to some extent, formative assessment, is also well advanced and in the case of the first, has considerable momentum through the Executive's engagement with it as it 'rolls out' across the Scottish educational system.

Where the picture seems incomplete or unsatisfactory is in regard to specific issues or particular strands within the general categories explored here. The concept of 'assessment burden' is a shibboleth employed with scant regard to evidence derived from untested assumptions. While many of these assumptions are tempered by insight and experience, there is a general lack of research into what the term actually means for learners and practitioners.

A number of critiques, many of which 'name and shame' assessment, have been levelled at Scottish education of late. Are they fair? Politically motivated? Reasonable? Given that Scottish education has largely developed through consensus and collaborative working throughout its long history, it would be in the interest of educational researchers to explore these criticisms and evaluate the evidence. The worst course would be to ignore it.

Something that seems to be missing altogether from research in Scotland is the impact of the international dimension of educational provision. It is expected that Higher National in China will prompt many new developments as a result of two different education systems attempting to 'talk' to each other. How does assessment change in such a context? Could, and should, SQA share its experiences of these international projects more widely?

There are also significant (and worrying) gaps with regard to race/ethnicity and to a lesser extent, disability. These are issues that are not going to go away and SQA should anticipate the need for more knowledge in this area, rather than wait to be asked to investigate. For all the rhetoric, assessment research in Scotland is still conspicuously ‘white’. Gender research is a little more advanced, but needs to be brought more effectively into the mainstream and to go beyond simple comparisons of exam performance. A look at the bibliography will reveal that the bulk of recent research into the area in Scotland has been by a single researcher, who has since left the field.

Exam performance itself, particularly international comparisons, is an area that will require further work; partly because of ministerial interest in how Scotland compares to other systems, and partly because there is something to be learned from other countries. There is also the sense that there are deeper questions to the issue than are currently being explored.

Above all, this is a good time to be considering research possibilities; 3–14, Portfolio Review, Assessment is for Learning, and the recently announced consultations extend the opportunity to SQA to investigate the detail behind the vision, provide the evidence to lead the practice, and give coherence to policy. Attention should also be paid to English and UK research projects and initiatives; although the UK is the stated focus of many of these, historically. Scotland is often assumed, rather than shown, to have the same characteristics as the rest of the country.

Recommendations

The most useful recommendation that can be made at this point is to study this review in detail, identify further sources from the bibliographies, and come to some agreement as to what seems most important to SQA. As SQA is currently developing a research agenda of its own, it would seem pertinent to include within this a dedicated assessment strand with defined research priorities, presumably tied in with those of the Scottish and UK governments.

Another dimension is European — colleagues in the continent are looking closely at the SCQF in developing their own frameworks, and there are ambitions towards a unified system of tertiary education. But what are the realities behind this? How compatible are our assessment systems? There is an opportunity here for SQA to be proactive in working with European colleagues to explore the issue.

There are many possible recommendations stemming from this review. There follows, however, some tentative recommendations as to areas that might be explored, perhaps as part of a proposed research agenda:

- ◆ The creation of a research strand, based on Scottish Executive initiatives, that will examine the role of assessment in delivering goals and targets. A research programme based on 3–14 and Portfolio would be a natural first step.

- ◆ Integrating assessment research into other areas of education research — is it really that separate from retention? Teaching and learning?
- ◆ Mirroring/duplicating UK and international research in a Scottish context. The ESRC TLRP programme is one such UK programme where SQA or partners could undertake supplemental research to explore the specifics of emergent findings.
- ◆ Revisiting our basic assumptions. Mhairi McAlpine’s summary of the different base assumptions that distinguish Scottish and English assessment methods explains much of why the systems differ. Are there grounds for a comparative study of the performance of a criteria-based and a norm-based assessment system?
- ◆ Researching SQA’s international projects, and the international dimension more generally, especially through comparative research. We know how Scotland performs in tests compared to other countries — but what are the systemic differences and cultural factors that affect these? Is it really as simple as saying that children in Singapore are more disciplined, or are there other factors at work? And what of the moves to build a joint credit and qualifications framework?
- ◆ Developing a research project into the realities of assessment burden.
- ◆ Dedicating some resources to researching the relationship between assessment and race, ethnicity, socio-economic factors, and gender.