

Assessment in Schools Fit for purpose?

Professor Mary James

University of Cambridge Faculty of Education

mej1002@cam.ac.uk



A new commentary on assessment policy

Authored by

Warwick Mansell, freelance journalist

Mary James, ARG and TLRP

Other members of the ARG:

Jo-Anne Baird

Paul Black

Richard Daugherty

Kathryn Ecclestone

John Gardner

Wynne Harlen

Louise Hayward

Paul Newton

Gordon Stobart



ARG

assessment reform group

T·L·R·P
TEACHING
& LEARNING
RESEARCH
PROGRAMME

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

A need for clarity

This commentary is an attempt to clarify some of the debates, bringing to bear the findings from extensive research, especially research from the Teaching and Learning Research Programme (TLRP) and the Assessment Reform Group (ARG), and to argue that this evidence needs to inform any future developments in assessment policy.

Hyperactive policy making

In England alone, at present:

- Reformed A Levels and GCSEs
- Introduction of functional skills tests
- New work-related and general vocational qualifications
- Uncertainty over testing at KS2 and KS3
- Assessing Pupils' Progress
- Piloting of Single Level Tests
- Assessment for Learning Strategy
- Proposed School Report Cards

A matter of purpose

Assessment has been asked to perform an increasing number of functions in recent years: from judging individual pupils to evaluating schools and monitoring national performance.

A key question for research has been whether it is effectively meeting all of the goals that have been set for it, and indeed, whether it is possible for it ever truly to fulfill so many aims simultaneously.

A distinction

It is helpful to make a distinction between the **intended use**, or uses, of assessment data, and their **actual uses**.

Results that are fit to be used for one particular (intended) purpose may not be fit to be used for another,

How many purposes?

Paul Newton has identified 22 different purposes of assessment. Each can comprise many more uses, e.g:

“Institutional Monitoring” can be used for: school-by-school performance tables; judgments on whether or not schools have hit test and examination targets; a basis for Ofsted inspections; performance pay and performance appraisal judgments for teachers and school leaders; assessments of teachers’ qualifications for promotion; within-school comparisons of the relative performance of different teachers; judgments as to whether schools qualify for school-by-school schemes including specialist school and training school status; appraisals as to whether they should be threatened with closure; and decisions on whether or not private companies running some schools and local authorities should qualify for incentive payments.

ARG

assessment reform group

T·L·R·P
TEACHING
& LEARNING
RESEARCH
PROGRAMME

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

The central problem

Assessment information has become a proxy measure that is supposed to facilitate judgments on the quality of most elements of our education system: its **teachers, head teachers, schools, support services, local authorities** and even the **government** itself.

This represents a fundamental change from the situation even 20 years ago, when test and examination results were predominantly meant to serve as indicators of what a **pupil** knew and understood of a subject.

Negative consequences

There may be negative consequences for the **pupil**, if an **institution** takes actions designed to improve its performance in the measured assessments which go against the young person's long-term educational needs, for instance, where teachers drill pupils in techniques for earning marks at the expense of teaching for deeper understanding.

Not a ‘no brainer’

It seems attractive, simple and cost-effective to use data from single assessments for multiple purposes.

David Bell, permanent secretary at the DCSF, told MPs in 2008: “While I hear the argument that is often put about multiple purposes of testing and assessment, I do not think that it is problematic to expect tests and assessments to do different things”.

We argue that clarity about the legitimacy of uses of assessment data, and the intended or unintended consequences of those uses, is crucial because the assessment system both **provides information** and **influences what people do**.

Three broad categories of uses

1. The use of assessment to help build pupils' understanding, within day-to-day lessons.
2. The use of assessment to provide information on pupils' achievements to those on the outside of the pupil-teacher relationship: to parents, further and higher education institutions and employers.
3. The use of assessment data to hold individuals and institutions to account.

Formative vs summative

“Formative” and “summative” are not labels for different types or forms of assessment but describe how assessments are used.

Assessments in class can be frequent mini-summative if they are simply used to establish where pupils are on levels and sub-levels, and to give them targets, without helping them to know how to achieve them.

Conversely, formal tests and exams can be used formatively if teachers help pupils to analyse their performance and find ways to improve their learning.

Assessment **quality** is a key
to judging fitness for purpose.

ARG

assessment reform group

T·L·R·P
TEACHING
& LEARNING
RESEARCH
PROGRAMME

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Quality in formative assessment(AfL)

- Part of pedagogy
- Responds to what pupils reveal of their (mis)understanding in lessons and helps them to move forward in learning
- Requires teacher expertise and peer support
- Aims to promote learning autonomy

Quality in summative assessment

Reliability and validity are central in all types of summative assessment made by teachers. Reliability is about the extent to which an assessment can be trusted to give consistent information on a pupil's progress; validity is about whether the assessment measures all that it might be felt important to measure.

Quality in summative assessment: in school use

- Teachers can sample the range of a pupil's work more fully than any external test.
- Initial mis-judgements can be put right quickly.
- But teachers need to think carefully about the validity of the assessments they use.
- And they need to resist the assessment tail wagging the learning dog.

Quality in summative assessment: by teachers for external purposes

Systems are needed to ensure that all teachers engaged in making judgments are working in comparable ways to an agreed set of criteria and standards e.g. intra-school moderation followed by inter-school moderation.

(Conflicts of interest can arise if the external purposes are high stakes for teachers.)

Quality in summative assessment: externally marked tests and exams

- Assessment needs to measure what it claims to measure, and to measure what is important.
- Need for cautious interpretation of results, taking account likely scale of measurement error.
- Should not undermine good teaching.

Quality in accountability: what are the data attempting to measure?

- The public routinely draws inferences from assessment results (about school quality) that go well beyond the inferences that the tests are actually designed to support (about pupils' attainments).
- 'Health warnings' are needed about the lack of information on context, measurement error, within school differences etc.

Quality in accountability: what are the consequences of publication of data?

- Accountability systems can damage the very outcomes they were designed to improve (by narrowing the curriculum, teaching to the test, choosing soft options).
- Governments need to trial new systems properly and monitor the consequences.

Quality in accountability: from micro-management to proper governance

“The pursuit of ever more perfect accountability provides citizens and consumers, patients and parents with more information, more comparisons, more complaints systems; but it also builds a culture of suspicion, low morale and may ultimately lead to professional cynicism, and then we would have grounds for public mistrust. In contrast, intelligent accountability concentrates on good governance and an obligation to tell the truth. I think [Parliament] has to fantasise much less about Herculean micro-management by means of performance indicators or total transparency. If we want a culture of public service, professionals and public servants must in the end be free to serve the public rather than their paymasters.”

Onora O’Neill, 2002, Reith Lecture 3

Four pressing challenges for policy

- Putting effective in-class assessment into practice system wide
- Enhancing confidence in tests and examinations
- Justifying the costs of assessment
- Avoiding micro-management

Putting effective in-class assessment into practice system wide

Values-practice gaps persist.

Necessary conditions for effective practice :

- Encourage a move from a performance orientation to a learning orientation
- Reduce time pressures (of curriculum coverage)
- Get rid of the tick box culture
- Support spread of knowledge and practice through teachers' collaborative classroom-focused inquiry (CPD)
- Give more priority to assessment literacy in ITE.

Enhancing confidence in tests and examinations

Tests and exams have far-reaching consequences – some unintended.

So:

- Treat results with due caution and be more transparent about the limits to the inferences that can be drawn
- Do more to enhance validity and reliability, based on sound research
- Consider the consequences on pupil motivation, teacher morale and school strategies of excessive ‘high stakes’.

Justifying the costs of assessment

In England, the annual direct and indirect costs of the assessment system is calculated as £750,000,000. Is this well spent?

Two points:

1. External assessments are not necessarily more expensive than alternatives, although they may not have the benefits
2. Pupils in England experience a huge load of formal assessment, from the early years profile to A Level

Avoiding micro-management

In England, the QCA – now Ofqual and the QCDA – has official control of the detail of how pupils are assessed. In practice, ministers have exercised extensive powers, partly through letters of remit to QCA requiring it to develop particular assessment changes.

Politicians have the right to determine overall assessment policy, but their involvement in specifying technical details of assessment models and procedures raises questions over whether they are sufficiently qualified to do so.

(Examples: Single Level Tests and functional skills tests)

Enhancing public understanding

The Commentary is an argument for the public to be alert to the far-reaching consequences of assessments and to ask hard questions about their fitness for purpose. The policies must, in the end, serve to advance the education of young people, not hinder it. In this sense, all assessment should be **for** learning.

Want to know more?

www.tlrp.org

www.assessment-reform-group.org

www.educationbynumbers.org.uk

