

First published 2004
by RoutledgeFalmer
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by RoutledgeFalmer
29 West 35th Street, New York, NY 10001

This edition published in the Taylor & Francis e-Library, 2004.

RoutledgeFalmer is an imprint of the Taylor & Francis Group

© 2004 John Gilbert for editorial matter and selection

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the
British Library

Library of Congress Cataloging in Publication Data

A catalog record for this book has been requested

ISBN 0-203-46409-5 Master e-book ISBN

ISBN 0-203-47065-6 (Adobe eReader Format)

ISBN 0-415-32777-6 (hbk)

ISBN 0-415-32778-4 (pbk)

CHAPTER 10

PURPOSES FOR ASSESSMENT

P. Black

Testing: Friend or Foe? (1997) London: RoutledgeFalmer, pp. 24–36

Introduction

Assessments may be conducted to serve several different purposes. In this chapter, the main types of purpose will be discussed in three main sections, concerned respectively with the support of learning, with reporting the achievements of individuals, and with satisfying demands for public accountability. These three sections will be followed by consideration of the interactions, supportive or conflicting, between these purposes.

It is important to match the selection and use of assessment methods to the particular purpose which the assessment is meant to serve and a distinction has to be made at the outset between the purposes and the instruments and procedures that might be used. For example, the same test questions may be used for quite different purposes, and, conversely, a single purpose might be served by combining the results obtained from a range of different types of assessment. Furthermore, assessments may be carried out by many different agencies, from the teacher in the classroom to a committee mounting an international survey. [...]

Support of learning

Formative assessment

The importance of this function is expressed in the following passage from a USA author:

... the teacher has need of constant information about what the student knows and the strategies being used to process and comprehend new concepts ... By imbedding diagnostic instruction in instructional activities, teachers can preserve the integrity of assessment tasks (the wholeness of tasks and natural learning context) and protect instructional time that would otherwise be diverted to testing ... There is general agreement that external packaged tests will not solve the problem of what teachers need to know about student learning.

(Shepard, 1992)

The issue is also emphasized, in a different way, in the following extract, ... from the TGAT report:

Promoting children's learning is a principal aim of schools. Assessment lies at the heart of this process. It can provide a framework in which educational

objectives may be set, and pupils' progress charted and expressed. It can yield a basis for planning the next educational steps in response to children's needs. By facilitating dialogue between teachers, it can enhance professional skills and help the school as a whole to strengthen learning across the curriculum and throughout its age range. (DES, 1988)

Taken together, these extracts cover a number of different points about the uses of assessment to improve learning. Outstanding is the principle that feedback is essential to the conduct of effective teaching. No system, mechanical or social, can adjust and adapt as it performs its task without that frequent information about the operation of its system which is needed to modify the input in the light of the actual — rather than the intended or imagined — progress of the system. This is particularly important in any enterprise where the response may be very variable and unpredictable — which is surely true of most classrooms.

Effective teaching ought to vary in pace and style according to the needs of the learners. Teachers need to go slowly, or repeat what has been done, when difficulties of their pupils become apparent. They need also to differentiate their teaching as they collect evidence that some have grasped ideas and want to go ahead, whereas others are trapped in confusions so that they are unable to go ahead. Whilst some differentiation may be achieved by setting, which can be a matter for controversy, all agree that even within a group chosen on the basis of a particular band of achievements, there will be a wide range of understandings and of rates of progress. Teachers need sound information on which to base differentiation decisions, and insofar as remedial action can be taken, its efficacy also needs to be checked. Individualized learning schemes are set up on the extreme view that pupils may be so different that each has to work at his or her own pace. However, the outstanding characteristic of such schemes is that assessment of attainment becomes the keystone in determining the progress of each pupil's learning work.

Ideally, assessment should provide short term feedback so that obstacles can be identified and tackled. This is particularly important where the learning plan is such that progress with this week's work depends on a grasp of the ideas discussed last week. Such assessment is generally called **formative**. It is clear that formative assessment is the responsibility of the classroom teacher, but others, in the school or outside, can support such work by providing training and methods.

All formative assessment is to a degree diagnostic, and the term **diagnostic** assessment is also used. Although it is hard to distinguish diagnostic from formative by dictionary definitions, in practice the two are used with different emphases, even though their ranges of application overlap. Standardized tests, constructed and refined by experts, are often used in a school to help identify special and extreme learning difficulties. The need here is for expert precision because the learning problems might be both deep and general, so that the type of adjustment that can be made on the spot in a normal classroom cannot possibly meet the special need. Such tests are said to be diagnostic. Thus diagnostic assessment is an expert and detailed enquiry into underlying difficulties, and can lead to radical re-appraisal of a pupil's needs, whereas formative assessment is more superficial in assessing problems with particular classroom work, and can lead to short-term and local changes in the learning work of a pupil.

Curriculum, pedagogy and the formative function

Because formative assessment is intended as the feedback needed to make learning adaptive and thereby more effective, it cannot simply be added as an extra to an

existing, non-interactive, scheme of work. The feedback procedures, and more particularly their use in varying the teaching and learning programme, have to be built into the teaching plans, which thereby will become both more flexible and more complex. This is emphasized very strongly by Linn (1989):

... the design of tests useful for the instructional decisions made in the classroom requires an integration of testing and instruction. It also requires a clear conception of the curriculum, the goals, and the process of instruction. And it requires a theory of instruction and learning and a much better understanding of the cognitive processes of learners.

(p. 5)

This quotation opens up further issues. The teacher has to decide what to assess, and to interpret the pupil's work in terms of a need that might require attention. In one topic, the fact that a pupil has not yet grasped a particular idea might not matter, in that it will be encountered again later and until then that grasp will not be needed. In another topic, a particular understanding may be an essential basis for work in the immediate future, so that the pupils will be disabled unless the present difficulty can be dealt with now. Judgments about what can be left and about what has to be grasped now are rarely clear cut. Thus, the practice of formative assessment has to be informed by a model that is quite detailed, in that it has to provide some guidance about the ways in which a pupil might progress in learning, linked to a clear conception of the curriculum and its learning goals. The reactions against multiple choice testing [...] were in part driven by the realization that they could not give useful information in relationship to new models of learning.

A different issue is raised by the above quotation from the TGAT report in its use of the phrase 'by facilitating dialogue', which follows from the statement 'a framework in which ... and pupils' progress charted and expressed', and which leads to hopes for strengthening the whole work of the school. The formulation of a policy about learning, and interaction between teachers and others about the framing and implementation of such policy, requires a language in which learning aims and achievements can be discussed. Such language must be linked to concrete examples, both of pupils' achievements and of teachers' actions in learning, for without this it will be merely abstract and idealistic with little practical effect. Formative assessment should provide those examples, develop this language, and enrich and so give meaning to the shared understanding of ideas and policies about learning.

All of these arguments should make clear that the improvement of formative assessment is a complex enterprise, yet one which should be at the heart of any policy which aims to improve pupils' learning, and which will also be at the heart of the enterprise of classroom teaching. It would be optimistic to assume that teachers usually have sound information about the progress of their pupils' learning — the evidence is that this is far from being the case. Many practical issues also arise, concerned with the collection, analysis, recording and interpretation of data for formative assessment and with the adaptations of classroom practice that may have to ensue. [...]

Certification, progress and transfer

Summative assessment within a school

Whenever pupils move so that responsibility for their learning is transferred from one teacher to another, information should also be transferred so that the new teacher

can plan work and guide each pupil appropriately. The information needed for this purpose will depend on how the work in the new class is related to the old. If there is close continuity, then the formative information about recent progress and immediate needs that the old teacher would have needed is also needed by the new.

However, the new teacher might also need an overview of each pupil's recent achievements and progress in order to be able to anticipate the progress and needs of each, which might affect the organization of the new classroom and the distribution of learning tasks. If the new work is very different from the old, for example if it involves a fresh start in which the subject is tackled afresh at a more sophisticated level than hitherto, then here again the information needed is more general — an overview of the pupil's earlier achievement which might help predict capacity to profit from the new learning programme.

Thus, on transfer the need for **summative** assessment arises. The term implies an overview of previous learning. This could be obtained by an accumulation of evidence collected over time, or by test procedures applied at the end of the previous phase which covered the whole area of the previous learning. Beneath the key phrases here, 'accumulation' or 'covered', lies the problem of selecting that information which is most relevant for the summative purpose. The principles of selection might depend on one's beliefs about what matters for the next phase of learning in the subject (and might therefore differ between different curriculum subjects) and might also depend on the particular transition that is involved.

As argued above, for transfer between classes in the same school, with a high degree of continuity, the difference between assessment for the formative purpose and for the summative might be rather small. Indeed, within the work of one year group, a teacher might gather formative evidence and supplement it with a comprehensive test in order to review and to decide on structural changes — for example to change the grouping of pupils. This might be seen as a weak form of summative assessment, happening quite frequently, and there would be a close link, and a difference of degree rather than of kind, between the teacher's formative and summative work.

However, transfer between different stages of schooling and between different teachers imposes new requirements. If it is to be an effective communication, the assessment information has to be formulated with a structure and a language that reflects a shared understanding between those who are communicating. It will not be enough that the two teachers communicating are working to a common scheme of work. It will be necessary also that the information satisfies three further criteria:

- It has to be adequately detailed. To say that a pupil has a grade C in science may convey little: a profile which showed any variations in this grade between (say) practical investigative work, learning of facts, and tackling of numerical problems would be far more useful.
- The two teachers have to be working to common criteria for grading. Given a report by a pupil on a practical investigation, one of the teachers might have given it a B and the other a D because they worked to different criteria — one teacher giving priority to orderly structure and clarity of expression in a report, the other to the quality of scientific thinking revealed in the investigation design and in the interpretation of the results.
- There should also be a shared procedure for determining standards of grading. Two teachers could be using the same criteria but have different standards for interpretation of the conventional grades in terms of marks — a decision by one to give (say) a B grade might be seen by the other as far too generous.

Transfer between and out of schools

These problems are more difficult to overcome if a summative assessment is to help guide the transfer of pupils between institutions — for example, between primary and secondary school, or from secondary school to a sixth-form college. Communication of criteria and standards will be less informal, so more priority has to be given to clear and agreed documentation. In addition, for each of the examples mentioned, several different schools will be sending pupils to the same higher institution. Unless these schools are working to common schemes and produce assessments on a shared basis the summative assessment information will be worthless. Thus, either different schools have to cooperate closely and accept constraints on their freedom to work in their own ways, or curricula and assessment have to be formally established and operated by external agencies. This is one of many examples where each school's own desire for its freedom of manoeuvre has to be reconciled with constraints which can be justified by the broader needs of their pupils.

Transfer to a new type of school can place a pupil in a very different type of learning environment. It will then be more difficult to predict, on the basis of previous achievements, how that pupil will progress in the new school. Ideally, the success of primary schools in predicting secondary progress ought to be investigated by monitoring the secondary school progress, so that their summative assessments can be improved. A great deal of such analysis used to be carried out when an examination at 11 (the 'eleven-plus') was widely used in England and Wales to determine the type of secondary school to which a pupil was to be directed. This analysis was used to modify procedures for interpreting the test — and school assessment — data to maximize the efficiency of prediction.

Such considerations also bear on the large discontinuities involved when pupils move either out of school into employment, or to further and more advanced study, whether at ages 16, or 18, or on completion of an undergraduate degree course. Here there may be very little direct link between the learning work, which might be the subject for any assessment and the future needs. The issue is further complicated because these future needs may be very diverse, ranging over various types and levels of further education and over many different types of employment. It is clearly almost impossible to imagine that a single grade can give information useful for all these purposes.

A set of grades obtained over several subjects may be a better guide. If, within each subject, the results were a profile reflecting success in meeting different types of demand, the user would have an even better chance of looking selectively at what was needed for his purpose. It might help to add to this information quite different assessments cutting across the boundaries of school subjects; possible assessments of so-called 'core skills' and assessments of personality characteristics are possible additions here. Overall, the need is for multi-dimensional data, but whilst such complexity can guard against simplistic judgments, it can also make demands on those using the information to study and understand its structure and its terminology, and also to keep up with the inevitable frequent changes.

Within this context, public examinations can be considered in relation to their function of certifying individuals. As will be discussed below, such examinations also have a function in judging teachers and schools for accountability purposes. In the UK, the national curriculum tests for ages 7, 11 and 14, the General Certificate for Secondary Education (GCSE); the Advanced-level (A-level), the National Vocational Qualifications (NVQ) and many others for older school pupils, and university degree examinations, are all in this category. These examinations are clearly summative, but

they may also contribute to teachers' formative work, by helping each teacher to check standards and by providing good examples of assessment tasks. In principle, such certification systems can use teacher assessments as part of their evidence. In some countries, the responsibilities are left entirely in the hands of teachers, in others teachers' judgments play no part. In most countries, external and school based assessments are combined in a variety of ways. [. . .]

Overall issues

The certification purpose of assessment raises three main issues:

- The assessment has to be as effective as possible in relation to its purpose. Here, the precise requirements can differ a great deal across this category — from transfer between infants and juniors at age 7 to the transition to employment of a university student on graduation. Each transfer requires its own methods, and what is generally lacking is research data, on the relationship between assessment results and subsequent attainments, which could be used to improve those methods.
- There have to be choices about who is to control the process and who is to carry it out. Here are included choices about the local or national control and about whether this control is general and flexible or close and detailed. Closely linked to such choices is the assignment of responsibility to teachers and schools, or to external agencies, or to a combination of these.
- Decisions have to take account of the costs implied. The cost of public examinations is a significant item in the budgets of secondary schools, and for all schools summative assessments which have to work to shared external standards bear a high cost in terms of the time of classroom teachers and of school managements.

Accountability

Accountability and test results

Schools have a responsibility to the public who fund them. One aspect of the discharge of that responsibility is to provide evidence that they are indeed promoting the learning of their pupils. An economical way to do this is to provide aggregation of the results of their pupils' performance in the various public examination systems which their pupils have to take for their own personal needs — i.e. those discussed in the previous section.

Assessment data will have to be detailed if they are to be a useful guide. However, such data on their own cannot provide valid guidance. A school working with pupils of poor educational background in an area of social deprivation where parental support is limited may be producing results which are below a national average but which, in the circumstances of that school, represent an outstanding achievement. Thus, data on the achievements on intake of the pupils, and on the catchment area and the pupils' home backgrounds, will also have to be considered in making any judgment or taking any action on assessment results. More generally, a wide range of data are needed if the achievements of the pupils in any one school are to be interpreted.

This is not to say that poor outcomes and real needs are to be hidden by some form of manipulation. Distinctions have to be made here — again in relation to the purpose for which the information is to be used. If the poor results of a school can be shown to be related to a low budget and to large class sizes, the policy guidance

is that action has to be taken on the resource determinants. If the performance and careers of the teachers in that school are to be judged, then allowance has to be made in the light of the factors which constrain them. If a parent has to decide whether or not to send her child to that school, then the judgment involves different considerations again. An alternative school may produce much better results, but if it also has a very favourable intake, it could be that for a given pupil of average attainment, that school may not do as well for that pupil as one which appears to have less impressive test results.

There has been much controversy about whether or not it is fair to publish school test results in 'league tables'. The justification that they can guide parental choice is a weak one, insofar as comparisons made without allowance for the many features which affect the results can be misleading and so lead to inappropriate choices. Many have argued that it would be fairer to schools and more helpful to parents to publish 'value-added' data, comparing intake test results with later attainments. However, a school's circumstances can affect its pupils' capacity to make progress as well as their starting points. Such complications will not be pursued further here — the general problems of judging school effectiveness raise issues beyond the scope of this book (see Gray, 1996).

Helping schools to improve

It may be natural to think of comparisons, but the public interest ought also to focus on whether certain agreed standards have been reached. Here the question of uniformity of assessments across schools becomes important. The detail and multidimensionality of reporting will also matter if the assessment is to indicate where the particular strengths and weaknesses of each school may lie. This last point raises the larger question of how those responsible for schools may act on the information about their work that assessment results may provide.

However, this question implies two assumptions. The first is that the assessment data are suitable for this purpose. For example, if a particular set of primary schools adjust their summative reporting to meet needs agreed with a particular secondary school, these reports may not be comparable with those in another district, even in the same urban area, so that comparisons cannot be made. Similarly, where the demands of local employment are quite specific, the adjustment of information to meet the needs of the local community may be in tension with a demand for uniform standards and for comparison with other schools.

Problems of interpreting results for policy purposes can arise within a school. Most secondary schools undertake careful examination of GCSE and A-level results. Such reviews can be helped by data on the average performances over the country in each particular subject and, where entry to a given subject examination is limited because it is a matter for optional choice, by comparisons of the general attainment of those pupils who enter for that particular subject. Because there can be quite marked differences between the entry groups for different subjects — some, for example, attracting the more capable, or repelling all but the most capable and committed, comparisons of performances between one subject and another have to be made with care, whether they are to apply at the level of school departments or nationally (see Fitzgibbon, 1996). Comparisons in relation to national assessment can be helped again by national data, both for test results as a whole and at the level of individual items. (Software for this purpose has been produced by the UK School Curriculum Assessment Authority.)

Sampling surveys

Another way in which assessments which might inform policy can differ from those for individual certification lies in the possibilities of sampling. To produce an overall picture of a nation's performance, it is not necessary to test every pupil within a given age group. The APU in the UK, for example, tested only about 2 per cent of each of three age groups. Moreover, even that gave sample sizes of about 12 000 for each test, whereas for any one set of questions it could be shown that a sample of about 500, if carefully distributed over different types of school and different regions, could give an adequately stable average. What was in fact done was to give about thirty different sets of questions to different selected sets of pupils. This made it possible for the surveys to explore a wider range of attainments in greater detail than would be possible with a test which would have to be the same for every pupil. In consequence, the data obtained were far richer, far more interesting and useful to teachers, and have been more relevant to policy debates than public examination data have ever been, or than the national assessment test data are proving to be (Black, 1990; Gipps and Stobart, 1993, pp. 34–36, 40–42). This approach, of giving different tests to different and restricted samples to enhance the value of the data, is known as **matrix light sampling**. It has been used in several countries, notably the USA, Australia and New Zealand, as well as in the UK's APU surveys. The issues here bear on the content and construct validity of assessment results. [. . .]

If the accountability to be secured is that of individual schools, it might still be possible to produce information that is less costly and more informative by giving different tests to different samples of pupils rather than by giving them all the same test. However, this could only be the case for large schools where any one age cohort would be large enough for the purpose.

Interactions between purposes

The choices which will distinguish or adapt assessments to their purposes may be made at several levels:

- At the outset, the particular set of test items or procedures may be chosen to fit to the purpose.
- Then the way in which these are administered and marked may differ according to the purpose.
- The ways in which the outcomes may be analysed or combined could also vary with the purpose.
- Finally, a particular purpose may require its own specific interpretation of that assessment data.

The three elements to be addressed here are assessment methods, assessment agencies and the purposes. One testing method carried out by a single agency might serve more than one purpose and would thereby be economical. If on the other hand, the methods to be used for different purposes have to be completely different, or the interests of different agencies are in tension or even opposed, then separation is required.

Certification with accountability

One example is the possibility that tests given to all pupils for the certification purpose could provide all that is needed for the accountability purpose. Most of the arguments which bear on this possibility have been touched upon in the previous section in the

discussion of sampling surveys. Sample surveys can provide better quality information at lower costs than the certification tests for every pupil, but if the certification test for all is a priority and has to be used, then the sampling survey comes as extra expense. The cost is not only financial. External assessments create work for teachers whilst making it more difficult for them to get on with their normal tasks. Thus, it was understandable, although regrettable, that with the coming of national testing for all at 7, 11, and 14 in the UK, the national surveys of the APU were discontinued in spite of the loss of the detailed information that the surveys alone could provide. Here it can be seen that the two purposes are in tension, so that to give priority to the one is to reduce a system's capacity to serve the other.

Formative with summative

A second and more notable example is the possibility that assessment by teachers might serve both the formative and summative purposes for their pupils and so remove the need for operation of separate agencies and procedures to serve the certification purpose. Some have laid stress on the differences between the formative and summative purposes, and have argued that the assessment instruments and procedures needed for the one are so different from those for the other that neither can flourish without clear separation. On the other side, it can be argued that the two functions are two ends of the same spectrum and that there is no sharp difference, and that if the two functions are separated, then teachers' assessment work will be devalued. [. . .]

Purposes in tension?

More will be said about these tensions in later chapters, for they raise more extensive technical arguments, whilst also moving into issues which are matters of public and political priority. What should be clear is that, in this second example as in the first, the two purposes are in tension. The time and effort needed by teachers if they were to bear the whole of the certification role would be extensive and might make it harder for them to develop and implement improved formative assessment. On the other hand, where very important decisions are based on wholly external tests, both pupils and teachers have to direct their work to meet the narrow range of demands which economical external testing can provide — and the model of teachers' own assessment inevitably becomes one of using examples of the external tests to train pupils, thereby weakening teachers' own formative assessment practices.

Whilst some degree of tension is inevitable, there can also be synergy. Instruments developed and trialled carefully by experts for certification and accountability exercises can be used by teachers to enrich their own range of questions used for the formative work. The work a teacher might have to do with peers to ensure common external standards when contributing to a certification process directed by an outside agency might well help that teacher towards a better appreciation of the aims and standards to which she should be working in her own formative assessment.

Summary

- The three main purposes of assessment are
 - formative, to aid learning,
 - summative for review, transfer and certification, and
 - summative for accountability to the public.

- The practice of formative assessment must be closely integrated with curriculum and pedagogy and is central to good quality teaching.
- The formative and summative labels describe two ends of a spectrum of practice in school-based assessment rather than two isolated and completely different functions.
- There are different levels of summative activity; summing up may be needed:
 - during the progress of work in one class,
 - on transfer between classes in a school,
 - for transfer between two schools, or between school and employment or further and higher education.
- The results of assessment and testing for accountability should:
 - be related to common criteria and standards,
 - be linked with comprehensive and detailed data on the schools intake and context, for otherwise the data will be unfair and misleading,
 - be designed and communicated so that they can serve the improvement of schools,
- Sample surveys may be a more efficient way of informing policy for the improvement of learning than blanket testing of all pupils.
- There are tensions between the different purposes of assessment and testing, which are often difficult to resolve, and which involve choices of the best agencies to conduct assessments and of the optimum instruments and appropriate interpretations to serve each purpose.

Bibliography

- Airasian, P.W. (1991) *Classroom Assessment*, New York: McGraw Hill. Chapter 1.
- Gipps, C.V. (1994) *Beyond Testing: Towards a Theory of Educational Assessment*, London: Falmer. Chapter 3.
- Gipps, C. and Stobart, G. (1993) *Assessment. A Teachers' Guide to the Issues*, London: Hodder and Stoughton. Chapter 2.
- Salvia, J. and Ysseldyke, J.E. (1991) *Assessment*, Boston: Houghton Mifflin. Chapter 1.
- Stiggins, R.J. (1994) *Student-Centered Classroom Assessment*, New York: Merrill/Macmillan. Chapters 3 and 4.
- Wood, R. (1991) *Assessment and Testing*, Cambridge: Cambridge University Press. Chapters 8 (diagnostic assessment), 17 (aptitude testing) and 18 (personnel selection and assessment).

References

- Black, P.J. (1990) 'APU science – the past and the future', *School Science Review*, 72, 258, pp. 13–28,
- Department of Education and Science (DES) (1988) *National Curriculum: Task Group on Assessment and Testing: A Report*, London: Department of Education and Science.
- Fitzgibbon, C.T. (1996) *Monitoring Education: Indicators, Quality and Effectiveness*, London: Cassell.
- Gifford, B.R. and O'Connor, M.C. (eds) (1992) *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction*, Boston and Dordrecht: Kluwer.
- Gipps, C. and Stobart, G. (1993) *Assessment. A Teachers' Guide to the Issues*, London: Hodder and Stoughton.
- Goldstein, H. and Lewis, T. (eds) (1996) *Assessment: Problems, Developments and Statistical Issues*, Chichester and New York: John Wiley.
- Gray, J. (1996) 'The use of assessment to compare institutions', in Goldstein, H. and Lewis, T. (eds) *Assessment. Problems, Developments and Statistical Issues*, Chichester and New York: John Wiley, pp. 121–33.

- Linn, R.L. (1989) 'Current perspectives and future directions', in Linn, R.L. (ed.) *Educational Measurement*, 3rd edn., London: Collier Macmillan, pp. 1–10.
- Shephard, L.A. (1992) 'Commentary: What policy makers who mandate tests should know about the new psychology of intellectual ability and learning', in Gifford, B.R. and O'Connor, M.C. (eds) *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction*, Boston and Dordrecht: Kluwer, pp. 301–28.