

Part II

Spreading the word, widening the practice

4 What is innovative about teacher assessment?

John Gardner

This chapter begins the discussion about developing teacher assessment in schools. It cannot be taken as self-evident that teacher assessment is a good thing, that teachers and schools should develop their practice in this direction. There are questions to be asked about its value and whether there is credible evidence of this. Experience would suggest that the underlying philosophy of ensuring that all assessment is for the good of pupils' learning is not widely adopted, though the advances of assessment for learning (AfL) indicate a strengthening trend. Part of the challenge of increasing the integration of teacher assessment in classroom and school practices is to ensure the changes necessary are well planned from a strategic perspective. As discussed in the introductory chapter, this means that a school must address a development process involving seven processes that overlap and intertwine in a complex and progressive manner. In Chapter 1 (Figure 1.1), these were represented as: Innovation, Warrant, Dissemination, Professional Learning, Agency, Impact and Sustainable Development. This chapter considers the first two of these: what is meant by innovation in an assessment context, and what evidence gives any particular innovation a warrant that convinces teachers and others of its worth.

Introduction

Innovation in education is a concept that defies simple definition, dependent as it is on the context in which it arises and the wide variety of social dimensions that the change process involves. Early work on educational innovation tended to focus on curriculum reform, such as the ground-breaking developments of the Schools Council/Nuffield Humanities Curriculum Project (1969–1972), or on theories of the 'innovation decision' process as derived from the empirical work of Rogers (1962) on

72 DEVELOPING TEACHER ASSESSMENT

aspects of the US agricultural industry. Much of what was written then and since centred on the transformation from concept or idea to policy and practice. As a consequence, the education literature is not short of evaluations of the innovations, including a modest corpus relating to assessment, that have swept through the system in the last 20–30 years. This chapter looks specifically at the concept of innovation in assessment but acknowledges that it is rarely easy to concentrate exclusively on an innovative idea without consideration of processes such as dissemination, which is dealt with in Chapter 5. This process and its allied professional learning activities sponsors the sharing of ideas and experimentation and the perspectives of those who are newly engaged begin to modify the original ideas and practices as they make their own contribution. As Schön (1971) observed ‘... innovation does not by any means entirely antedate the diffusion process; it evolves significantly within that process’ (p. 107).

Innovation, even with high intrinsic value, may be of little interest if there is no attempt to transform it to routine practice, where ‘routine’ is not merely automatic but signifies that it is sufficiently well regarded and used, it constitutes commonly expected practice. However, existing practice cannot be transformed unless an appropriate innovation is brought to the fore. Consequently, this chapter focuses on the concept of innovation itself, while reflecting at some points on the processes that bring it to wider audiences and application, and ultimately to sustained implementation.

The nature of innovation

Sometimes in scientific or medical contexts, innovation may be almost serendipitous, arising from chance and good fortune. More often than not, however, it will be the result of years of painstaking research; for example, in some genetics contexts. In education, innovation is not as likely to be a discrete outcome of a research process or even a ‘... tidy picture of a coolly managed process’ as Ruddick (1976: 5) prefaced her report on the humanities project. Conceding that the report may have been misleading in this respect, she acknowledged that it missed the ‘... puzzlement and opportunism that characterize such ventures ... and the sense of responding to events rather than controlling them’; a picture more resonant of innovation by evolution than revolution, however dramatic the ultimate shifts in practice may appear to be.

Arguably, therefore, educational innovation emerges in a more ‘organic’ fashion. For example, it may follow a bottom-up variant of Rogers’s centre-periphery model in which a new idea emerges at the researcher/teacher interface, captures the interest of increasing numbers of teachers

WHAT IS INNOVATIVE ABOUT TEACHER ASSESSMENT? **73**

and grows from its small beginnings to eventual adoption by whole sectors of education. AfL and its espousal in the UK was just such an innovation that grew from the seminal review of research on formative assessment by Black and Wiliam (1998a) and from a subsequent, enthusiastic uptake of the key findings by teachers and schools. By way of contrast, the history of using computers in the classroom has been considerably more problematic. The introduction of computers has closely followed the top-down centre (government) to periphery (school) variant of Rogers's classical model, but has arguably not been effective in promoting classroom transformation despite the massive 'seeding' afforded it in successive waves of multi-million pound funding initiatives by government. As a major educational policy innovation, largely isolated from the operational context of schools, it is debatable whether it is viewed more as an end in itself than a means to improving learning. To all accounts this particular innovation continues to stutter. For example, the Office for Standards in Education (OFSTED, 2004) has reported that '... the government's aim for ICT to become embedded in the work of schools [is] a reality in only a small minority of schools ...' despite '... good evidence to suggest that most teachers regard ICT positively, with only a residual minority of the profession reluctant to take their work forward with ICT' (p. 6). Despite teachers accepting the use of computers in education as a 'good' thing, it has not been enough to initiate the deeper changes necessary for integrating it into practice, a phenomenon also identified by James and Pedder (2006) in AfL contexts.

The last few decades in education have also seen the rise of the gurus or evangelistic educationalists who purvey ideas and innovations with a charismatic panache. They take large numbers of their audience through the first two 'knowledge' and 'persuasion' stages of Rogers's (1983) five-stage innovation-decision process ('decision, implementation and confirmation' being the subsequent stages, p. 20). Thinking skills, multiple intelligences and AfL are all recent examples of inherently important educational innovations in educational practice or theory. However, they have also become vulnerable to criticism from traditionalists, primarily because they have been the subject of relatively trivial expositions by some 'true believers'. Replete with classroom anecdotes, the appeal of the guru's message to hard-pressed teachers is often sufficiently seductive for them to launch straight into a regime of tips and tricks. However, the effect can be very short-lived as the same teachers soon jettison them because the cycle of reflection and action has not been fully engaged and the deeper theoretical and philosophical assimilation of the innovation has been missed. As Fullan (1993) puts it: 'It is not enough to be exposed to new ideas. We have to know where new ideas fit, and we have to become skilled in them, not just like them' (p. 16).

74 DEVELOPING TEACHER ASSESSMENT

The increased importance of reflection for promoting improvement and innovation in all forms of successful professional practice is usually attributed to the work of Schön (1983). Although his examples are largely drawn from industry, they recount a symbiotic relationship, mediated by reflection, between theory on the one hand and empirical knowledge on the other. His outline of the invention of the junction transistor at Bell Labs (p. 181) charts a process in which ‘... reflection on theory leads to experiment... [then]... reflection on the unexpected results of experiment leads to theory, or to invention’. Shockley, the principal scientist involved in this example, reputedly called this method ‘creative failure methodology’ (compare this with Schön’s euphemistic ‘unexpected results’) but the essence of it was the experimentation and reflection on the results. When teachers try out new (to them) techniques, the hope is that they will also reflect on ‘what works’, thereby using their own experience to amend any accompanying theory that has been designed to explain the proposed effectiveness or application of the techniques. And just like Shockley’s experience, any innovation or invention in classroom assessment may be prone to initial ‘failure’ and reflective revision, sometimes humorously summed up as the ‘Ready! Fire! Aim!’ cycle.

Innovation in assessment, then, is often promoted at the level of the individual by encouraging the trialling of new methods, evaluating their effectiveness and worth, and modifying practice in the light of this reflection. The notion rarely holds that the worth of an innovation is self-evident or guaranteed by theory; implementation or trialling at some level is a clear requirement. Innovation in assessment needs sufficient visibility for the individual to be aware of it, know how it operates and what can be achieved. Ultimately teachers need to experiment with the innovation in order for them to begin to transform their own practice.

Innovation in assessment

In science and medicine, major innovations are often definably ‘new’ and discrete (e.g. a new drug) while in education they can be reincarnations of older practices, or new ways of carrying out established activities, all cast as innovative. What, then, do we understand innovation in assessment to be? The dictionary definition might lead us to expect a ‘new’ type of assessment, new in terms of the methods used or the process undertaken, or indeed the focus. The much quoted ‘new learning’ heralded by twenty-first-century technology and its impact on society arguably presents significant assessment challenges, the resolution of which may well constitute innovation. There are various candidates for this new learning. For example, QCA (2007a) launched a framework for learning and thinking

WHAT IS INNOVATIVE ABOUT TEACHER ASSESSMENT? 75

skills. This covers the need for learning and, by implication, assessment in areas such as: teamworking, independent enquiry, self-management, reflective learning, effective participation and creative thinking. A recent input by Baker (2007), the then President of the American Educational Research Association (AERA) has also heralded the need to assess adaptive problem-solving, risk assessment, managing distraction, self-management and changeable roles.

As a new focus for learning, any approach to appraising skills of 'distraction management' might, therefore, have a legitimate claim to being innovative. More often than not, however, the newness identified in innovative contexts is in fact 'situated' or context-dependent. In most situations, it is probably fair to say that the assessment process is not so much new *per se* as it is new to those people, those circumstances, those places, and so on.

Take for example formative assessment, assessment that is designed to support learning. The importance of this form of pedagogically integrated assessment has come to the fore in recent years. Based primarily on Black and Wiliam's (1998a) review of the research evidence supporting the effectiveness of formative assessment in promoting learning, ARG (2002b) launched its 10 principles to promote and guide the practice of AfL. AfL has been a force for change in classroom practice in national assessment systems in the four nations of the UK (see Chapter 3) and wider afield in, for example, the USA, Canada and Europe (OECD, 2005). It could therefore be perceived in some quarters as an assessment innovation that has swept across the global education landscape, fulfilling one of Mayer's (1991) indicators of successful innovation: the transfer of 'content, methods and actions . . . the creation of connections between "different disciplines . . . different teachers, their methods, their value systems and their behaviours"' (cited by Elliott, 1993: 60). A more grounded view might be that it merely constitutes good pedagogical practice being introduced in places in which more didactic practices had long held sway. Such 'places' might be as singular as a teacher's classroom or as cross-cutting as all schools in a particular local authority. They certainly include whole sectors such as higher education, where a sea-change in assessment approaches over the last decade has fostered such 'new to the sector' innovations as peer- and self-assessment, and criteria sharing (see, for example, Boud and Falchikov, 2007 and Bryan and Clegg, 2006).

Identifying innovation, therefore, is not a simple matter of perceiving a change that some consider to be novel. As this argument would imply, we must first identify whether it is widely recognized as genuinely novel and perhaps even experimental, never having been used before or used only in very localized and isolated circumstances. If this test fails, we must next determine if it is contextually new: new to primary classrooms in general,

76 DEVELOPING TEACHER ASSESSMENT

to a set of schools (such as a local authority), to a specific school or indeed to a particular group of staff or individual teachers.

Simply registering a change of practice as 'innovative', and then attempting to appeal to a person's curiosity and professional interest, can be damaging. Indeed, such an approach may well constrain further development if those who are the targets for adopting the innovation do not recognize its novelty aspect or, worse, see through and reject it, perceiving instead a top-down directive that is designed to promote some form of unilateral behaviour change. From another perspective, however, using words like 'innovation' can serve as the Trojan horse that avoids telling the target group that their current practice is inadequate!

Effective adoption of an innovation is widely considered to be dependent on its 'ownership' by those who must adopt it; ownership being more to do with personal beliefs, and the promise of self-benefit and benefit for their students, than mere changes in practice or behaviour. As Morrish (1976) put it: 'People generally accept innovations more readily if they understand them, regard them as relevant to their particular situation and also help to plan them' (p. 129).

Types of innovation in assessment

Given the caveats above, it is something of a tall order to identify the types of innovation that may be encountered in assessment contexts. Broadly speaking, they form a gradation in 'newness'. Beginning with arguably the least innovative and ending with the most, let us call these:

- innovations in administration (facilitating assessment processes, record-keeping and reporting);
- situated innovation (assessment practice that is new in the circumstances);
- innovations for 'new learning' (new aspects of assessment specifically addressing twenty-first-century goals).

Innovations in administration

Perhaps among the easiest 'innovations' to identify are those that relate to the administrative processes of assessment. Assessment carried out through the medium of computers, for example, is often misleadingly described as an innovation in assessment when it is more precisely viewed as an innovation in the administration of the assessment. Computer-based or online offerings stretch the concept of innovativeness if they are merely pen-and-paper tests presented on screen, with the examinee's responses typed directly into the system. There certainly was a time when

WHAT IS INNOVATIVE ABOUT TEACHER ASSESSMENT? 77

assessment through the medium of a machine was a new and frontier-pushing development (e.g. with Pressey's 1926 'simple apparatus that gives tests and scores') but the assessment *per se* was no different from that which could be carried out by a person; it was simply considered more efficient.

This is still the case today. The potential for cost savings and administrative efficiencies are aspects of standard online assessments that are regularly argued as selling points. However, they fail any reasonable test of innovativeness in assessment itself. Objectivity is another selling point but even here the concept is made manifest in a relatively minor way. The claimed objectivity may be based on applying indisputably correct answers in a process that could be carried out just as objectively by human judges (a simple example of the type of item might be the case of 4 being the indisputable answer to $2 + 2 = ?$). Or it might be based on the application of fixed answers, previously interpreted and supplied by human judges and therefore potentially subject to the human error they purport to defend against. It is therefore not quite the holy grail-like objectivity of no human error. Nor is it the 'objectivity' of finely honed subjective judgements that have been reviewed and endorsed by several to many human judges in a rigorous moderation process.

A significant step up from the test that is merely computer-based is what is known as an adaptive test, headlined by the in-vogue computerized adaptive tests (CATs) of recent years. In the case of CATs there is perhaps more justification to apply the term 'innovation in assessment'. Adaptive testing is a relatively dynamic form of assessment that proceeds in a cycle. The first assessment of the pupil identifies the level of difficulty in terms of items that they can manage. The next stage of the examination process is then tailored to a level of difficulty at and above the assessed capability level. Again, the level of difficulty the pupil can manage is reassessed. The test continues through these cycles of tailored assessment until the pupil can no longer 'master' the level of difficulty of the examination questions being presented.

In a paper version, the process would be very limited and self-directed; for example, 'If you have answered questions 5 to 7 correctly, please proceed to Section D, Question 11 ...', and so on. The considerable perseverance and honesty required of anyone taking a paper-based 'programmed' learning test rules it out in most cases but the computerized versions are considered to carry out assessment and capability levelling processes, such as attributing pupils' work to national curriculum levels more rapidly and objectively. The innovation lies in the development of sophisticated algorithms for calibrating a large collection of assessment items, establishing the level of difficulty at which a pupil is currently working and processing their pathway through the items at appropriately increasing levels of

78 DEVELOPING TEACHER ASSESSMENT

difficulty. The conclusion that proponents of CATs promote is that the last level the pupil can manage is an accurate measure of their achievement. Simple as it might sound, systems based on this approach have significant failings (see, for example, Way et al., 2006 and Wise and Kingsbury, 2000) but over time many of these are responding to increasing refinement and sophistication. Various claims are made about their ability to deal with more complex assessment contexts such as those presented by the creative and expressive learning domains (see, for example, Embretson, 2003) but in most existing cases CATs are best suited to multiple-choice and fixed answer designs.

Situated innovation

Part of the argument under this heading hints at a restricted type of innovation in assessment; namely, that which is new in the circumstances in which it is introduced or observed. Consider a classroom in which a history teacher reads from a chosen specialist text on an aspect of history that the students must study. The students listen and eventually the teacher closes with 'Any questions?', possibly in that perfunctory manner which signals that questions are not really expected or desired. The essay assignment is given; the end of class is signalled. A bit 1950-ish perhaps, but this type of learning experience is not exactly extinct. Clearly, it would be an innovation if, in another scenario, such a teacher engaged the students more directly, in debates, role plays, site visits, research tasks, project work, and so on or even more simply in genuine one-to-one, group or whole class discourse about the matters under study. Such pedagogical tools are known to be effective in promoting deeper learning and would be innovative in the circumstances of the classroom described. However, arguably, it would be a further innovative step if the teacher were also to integrate assessment formatively into the learning process through appropriate sharing of learning objectives and success criteria, questioning, feedback, self- and peer-assessment, and the identification of next steps to improve the assimilation of the learning.

The two 'improved' scenarios above could be described as arising from innovation in the teacher's approach; the one more specifically pedagogical; the other relating to the use of assessment to support learning. However, experience has shown (e.g. James et al., 2006; Leitch et al., 2006) that in some classrooms these 'innovations' may not be deeply assimilated into professional practice. Instead, they may be treated superficially as a set of teaching tips for improving student engagement and motivation to learn; innovative in the circumstances but not reaching the potential for which they are designed. As Fullan (1993: 23) comments: 'It is no denial of the potential worth of particular innovations to observe that unless deeper

WHAT IS INNOVATIVE ABOUT TEACHER ASSESSMENT? 79

change in thinking and skills occur there will be limited impact'. Arguably, of course, any means to improve students' engagement and motivation, however limited, are surely to be welcomed.

Whether treated superficially or deeply espoused, it is clear that formative techniques would not be considered innovative in circumstances in which teachers have an underpinning grasp of the importance of using assessment to support learning and already use some or all of them in their day-to-day practice.

Knowing when something is innovative with good, long-term impact and not merely a novelty that promotes short-term success, is not a widely held skill. For this reason, the need for a deep understanding of what may be a purposeful innovation cannot be underplayed. Too often teachers, the school system and policy makers are regaled with the latest ideas. Some of these are not much more than fads but have been cast in the 'must do' urgency of some of the less thoughtful voices in the school improvement lobby or indeed from policy makers with short-term political agendas.

Arguably, at the root of all calls for change today is the aspiration to improve the learning experience and outcomes for every student in every classroom, through the improved teaching and facilitation of learning by their teachers. The continuous cry of falling standards reverberates through the system whether aimed at national examinations (GCSEs, A-levels, etc.) or at basic skills (levels of literacy and numeracy, etc.). Yet many commentators and researchers reject the notion that the standards in use in education in England, for example, have other than a very limited value in appraising the quality of education. One of these commentators, Mansell (2007), contends that any mention of standards should come with a 'health warning' (p. 26). In his view, the public notion that *raising standards* means raising the quality of the education provided is seriously out of kilter with the reality in schools, and that the concept of *raising standards* is reified in many schools simply in the aim to raise test scores.

This phenomenon of the link between standards and examination outcomes continues despite the many deep-lying social issues that are known to mingle with the educational and pedagogical dimensions of schooling. Inevitably, it is schools that take the brunt of both the blame and the responsibility for rectifying what is in essence a misconceived issue. And the ensuing calls for change command an audience at the highest levels from government departments through their statutory curriculum and assessment agencies to local authorities. Action, often cloaked in the terminology of innovation, typically plays out through government consultations, pilot studies, professional development programmes and voluminous resources, often online, on CD/DVD disks or in glossy printed packages. None of these actions warrants criticism *per se* but they have the potential

80 DEVELOPING TEACHER ASSESSMENT

to suggest a patronizing ‘we know what is best and what you should do’ approach that seems to have missed the decades of ineffective impact on which those proposing such an approach could reflect. However, this is not to say that it is unnecessary to have good resources or evidence from well-carried out pilot studies; who would not benefit from having them when undertaking an innovative change to their practice?

Innovations for ‘new learning’

‘New learning’ is a term that is bandied around the education system, nationally and internationally, with only the most basic of commonly held understandings and no widely accepted definition. As more or less a sound-bite concept, it can attract an audience with policy makers, academics and teachers alike, much quicker than most educational issues. But it is likely that these various audiences hear different things. The ‘knowledge that is of most worth’ in today’s society could arguably draw on Herbert Spencer’s ‘science’, a continuously evolving adaptation to the modern world, probably more generally couched in terms of ‘new’ skills that are perceived as necessary. To policy makers it might be literacy, numeracy and ICT skills; to society more generally it may be aspects of citizenship, and to academics it might be thinking skills or skills to manage distraction. Whether the target knowledge, understanding or skill is actually new or simply in vogue, the question arises as to what form of assessment best addresses it.

Much of what is claimed to be innovative in assessment actually derives from considerations of how validity in the making of assessment judgements may be improved. For several decades there has been a rumbling unhappiness with psychometric and standardized testing programmes, the types of assessment that give rise to scores, marks and grades. The reputed high reliability and acknowledged high costs of external testing has also come under significant fire in the UK; the former because it is not always the case; the latter because it is largely unwarranted (for a brief summary of the positions on reliability and cost, see Gardner 2007). Such measures have little meaning in relation to the learning they have been used to assess and generally have even less prospect of contributing formatively to students’ learning. The consequences of their dominance, however, include ‘wash-back’ damage on other aspects of the system:

... increasing the use of externalized methods and reporting [it] has eroded trust in the professional judgement of education practitioners to deliver assessment in other contexts. Above all, it has had increasingly serious consequences for the system’s overall fitness-for purpose.

(Skidmore, 2003: 45).

WHAT IS INNOVATIVE ABOUT TEACHER ASSESSMENT? **81**

A variety of alternative or innovative methods of assessment have flourished over the same several decades, as a means of raising the ante on validity. All of them have relatively shaky histories in terms of adoption, owing partly to the continued hostility of some influential policy makers towards anything perceived as subjective assessments and partly to the logistics and costs of the moderation and validation of the judgements provided. The overarching innovations in question can be conveniently classified as coursework and authentic assessment.

Coursework covers a number of possible assessment vehicles, including portfolios, project work, exhibitions and oral presentations. It has had a chequered history and indeed has often suffered a lack of confidence (and therefore investment) in the teacher assessments used, the standards of work achieved or the fear (with some justification) of significant plagiarism or third-party support for any of the unsupervised aspects of the work. Recent moves in national examinations at GCSE level in England, Wales and Northern Ireland (QCA, 2008a) have witnessed the introduction of 'controlled assessment', which may address some of the perceived problems (see Chapter 3 for further information on controlled assessment).

The early 1990s also saw the emergence of the concept of 'authentic assessment', particularly in the USA. Initially promoted as a high-validity alternative to the perceived low validity of external and state or national testing, it aims to assess learning in a manner that relates more closely to the way in which the learning content arises from or affects students' daily lives. Authentic assessment therefore avoids psychometric or externally administered tests, using instead the same types of assessment approaches as are used in coursework; for example, portfolios, research-based projects, presentations and exhibitions. However, even the 'everyday life' innovation of authentic assessment seems to have lost its way in some quarters where it now exists primarily as a 'rubric-based' approach to integrating curriculum, performance standards and assessment.

A rubric, as the name suggests, prescribes the type of learning to be undertaken, the assessment criteria to be used and the standards of performance onto which the criteria map. Widely used in the USA, the example of Performance Standard 24B.E (Illinois 2007) illustrates the model. This health education rubric requires teachers to develop students' competence in applying their knowledge and exercising their decision-making skills in two out of four 'real' life scenarios provided. These paragraph scenarios describe an incident and the rubric identifies how the students' responses to a prescribed decision-making process should be graded. Electronically scanned examples of student work that either 'meets' or 'exceeds' the standards of performance are also provided to complete the all-encompassing nature of the rubric guidance. Assessment by teachers may be an innovative element of this approach but the dependence on pre-ordained rubrics

82 DEVELOPING TEACHER ASSESSMENT

arguably gives an up-to-date meaning to Dewey's (1938) counsel that: 'Nothing has brought pedagogical theory into greater disrepute than the belief that is identified with handing out to teachers recipes and models to be followed in teaching' (p. 170).

A major element of the debate about assessing new learning, though not always acknowledged or expressed, is that a focus on content is potentially no longer valid. If a curriculum aspires to develop autonomy or self-reflection, one question for the assessment community might be: Can these be assessed without recourse to content-based proxies? How valid is the assessment made of a student's 'ability to work as a member of a team' when it is based on a process that attempts to disaggregate the individual contribution from that of collective endeavour in a group project? At essence, the central question arising in relation to assessing new learning might be conceived as: Is this judgement of what a person knows/has learned/understands, and so on valid in terms of the evidence used to make it and the process used to collect the evidence?

Green (1998) counsels us to remember that 'judgement' has at least two meanings. In an assessment context, the first would be the process of assessing and deciding the level of achievement and quality of a student's work and the second would be the category decision itself (i.e. a grade, level or score). He argues that such judgements are never merely subjective (e.g. whimsical or unsupported by evidence) as they are always based on '... reasons, grounds, rules or principles' (p. 178) or, as might otherwise be argued, on the evidence available against commonly held standards and level descriptions. However, it is entirely possible, and not uncommon, for two assessors to interpret the same evidence and arrive at a different judgement or for two assessors to examine different types of evidence and arrive at different judgements about the same performance. In the complex scenarios of new learning, such challenges are writ large.

Innovative types of evidence or means for collating it, which are designed to increase the validity of evidence on which assessments are to be made, are only part of the story. Take, for example, the humble UK driving test. Not so long ago the 'knowledge' part of the test was carried out through Highway Code questions, which were randomly chosen by the examiner and presented orally at the end of the practical driving test. The aspiring driver could be 'failed' for answering a question incorrectly, even if the practical aspects had been exemplary. More recently the examination involves a computer-delivered 'knowledge' test, the passing of which, at a preset threshold, determines whether the student driver proceeds to the practical test. This test enables their competence in practical driving skills to be assessed, with the examiner's assessments largely governed by preordained competence thresholds. However, the examiner retains a degree of discretion over elements relating to the student's control of the

WHAT IS INNOVATIVE ABOUT TEACHER ASSESSMENT? **83**

vehicle, smooth use of gears and brakes, and so on. What then might constitute innovation in this assessment setting?

Validity is clearly an issue. Yes, the computer-based test is an efficient, cost-effective and objective means of testing certain types of knowledge relating to driving. And, yes, the practical test is a relatively valid means of accessing actual competence. But it is a test of only 50–60 minutes, duration, which cannot cover all possible driving situations, manoeuvres and skills. In assessment terminology, the validity of this test, in common with the large majority of tests in any context, is challenged by the restricted learning domain that it is able to assess in the time and circumstances. What might be innovative in this context, therefore, would be the keeping of a log of the driving experiences as the student driver is learning how to drive. This could serve as evidence for accredited instructors to vouch for the competence of their student drivers, when they judge it appropriate. The analogy could extend to the obvious challenges such a ‘more valid’ system might present, but there is general wariness about the dependability of tutors’ judgements of their own students’ performance in many contexts, not least when the judges are teachers in schools. However, if these judgements are made in constrained circumstances for the various types of ‘new’ learning under scrutiny, all the existing challenges to validity (and reliability) will likely persist and become even more vulnerable to negative critical scrutiny.

Conclusion

Since the 1960s and 1970s, the psychometric grip of the psychologists on assessment in schools (and elsewhere in education) has been progressively challenged in relation to the unwarranted claims of reliability in many instances and to a lack of validity in most. In parallel with these challenges, there has been a rising demand for meaningful assessments, which in turn has given rise to a plethora of innovative approaches to assessment, variously hailed as authentic, valid and purposeful. Today, alternative and ‘innovative’ approaches to assessment in schools include portfolios, project work and presentations. But such innovations in assessment may not be all that they seem; indeed, they may not be innovations at all.

Central to all of them is the practice of assessment and judgement-making by teachers, for both summative and formative purposes. What has been argued in this chapter is that in many respects, the innovative dimension of some approaches may not actually be an innovation in assessment; it may be more of an innovation in assessment administration or a situated assessment innovation that is ‘new’ to the teacher, to the

84 DEVELOPING TEACHER ASSESSMENT

school or to the circumstances in which it is introduced. Alternatively, it may be innovative in assessment, in for example, striving to address important and currently unfulfilled assessment needs, such as those demanded by the curricular and pedagogical pursuits of 'new learning. It is important therefore to analyse an innovation to determine whether it is actually innovative and, if so, in what circumstances and why it is considered innovative. Once determined, these will contribute to a better understanding of what is being proposed as an innovative change in educational practice. This will in turn contribute to considering how effective it has been, what value it may hold for the target audience and how best it may be transformed into well-regarded and common practice.

Questions for reflection

1. What would be the features that would identify a certain practice of teacher assessment as an innovation rather than a part of regular practice?
2. What information and experiences with an innovation in teacher assessment would be required to prompt others to integrate it into their practice?
3. How might innovations be introduced and developed to promote changes in both understanding and practice?
4. Reflect on a recent assessment innovation. To what extent was this introduced and developed in ways consistent with ideas in this chapter?