

Chapter 12

ASSESSING PERFORMANCE

Why you might find this chapter interesting

*It has been our experience that the better we are able to assess learners' performance, the richer can be our approaches to developing it. This chapter explores this tight interrelationship of teaching, learning and assessment. We begin by exploring the **purposes** of assessment, and what happens when there is confusion or disagreement about those purposes. We reassert the centrality of **performance** assessment, designed to provide insights into the **capability** of learners. This capability however is not a single monolithic quality, but can be flavoured to emphasise particular perspectives on capability. We discuss frameworks for creating authentic assessment tasks and techniques we have used for presenting these to learners in convincing ways. We explore the substructuring of activities and the critical role that this plays in revealing evidence of learners' capability.*

Once the activity has been completed, the challenge changes to one of making judgements about the quality of learners' work. We discuss the role of (and the relationship between) holistic and atomistic approaches, of rubrics and approaches to using them effectively, and explore the challenge of assessing group performance. We then outline a radically different approach to assessment (differentiated pairs) that does not involve any 'marking' or attributing of scores, but is based rather on multiple direct comparisons of pieces of work. It is an approach that demands performance in digital form (web portfolios) and our use of it is currently attracting great interest from assessment and policy bodies. We conclude the chapter with some reflections on our whole unpickled portfolio approach to assessment.

Learning and **teaching** processes are two of the key threads running through so much of our work and in both cases have consistently been focused on how we might better understand and promote design & technological **capability**. Inevitably, therefore, we have also continually been drawn to the challenge of assessment – for to **understand** capability and to have a sense of how one might **develop** it, one also needs to be able to get a grip on it – to weigh it – to judge its constituents and its quality. It has been our experience that the better we are able to assess learners' performance, the richer can be our approaches to developing it.

1. THE CENTRALITY OF PURPOSE

The starting point for considering assessment must be that of purpose, as the nature of the purpose will have profound consequences for the form of assessment that might be most efficacious. Nonetheless, we recognise that there are many examples of systems designed for one purpose, which are then forced into wider service for additional purposes. This multipurposing frequently has damaging repercussions since tools designed for one purpose seldom operate as well when used for a different one. The blood that is frequently spilt when screwdrivers (and even chisels) are used as paint-can openers provides a salutary lesson. In the education world, far too much blood has been spilt using National Curriculum Standard Assessment Task scores (summative motive) for constructing league tables of schools (evaluative motive) – something we witnessed first hand through our involvement with the initial Standard Assessment Task development projects.

This is not to say that a particular assessment tool cannot be exploited for different purposes where the context and assessment intentions can be well serviced by the use of the tool. To illustrate both the problems of purpose and the potential of a multipurpose assessment tool, *APU Design & Technology* provides a useful example. The project was principally an **evaluative** project, designed to enable us to report on the performance of 15-year-old learners in England, Wales and Northern Ireland.

The strength of this light-sampling approach was that it enabled us to develop an extremely comprehensive set of tests that, taken together, reflected a complete perspective on capability in design & technology as described at that time. Because the learner sample was so large, we were able to distribute the tests across the sample, so that no individual learner undertook more than two, i.e. 3 hours of testing. If each learner had taken the whole test battery, it would have taken them 36 hours to complete. But *APU* procedures enabled us to construct a picture of overall national performance by amalgamating the performance of so many individuals, each doing their bit.