

Contextualising Authentic Assessment

Assessment in Education: Principles, Policies and Practices (1999) 6(2), 177-194.

J. Joy Cumming

*School of Cognition, Language and Special Education,
Faculty of Education, Griffith University,
Nathan 4111, Australia.*

Graham S. Maxwell

*Assessment and Evaluation Research Unit,
Graduate School of Education, The University of Queensland,
Brisbane 4072, Australia.*

September 1997

ABSTRACT *The term 'authentic assessment' has recently gained widespread use in education. This paper explores various ways in which authentic assessment is being interpreted and the relationship between these different interpretations and the original focus of authenticity in learning. The paper explores briefly the ways in which implicit and explicit beliefs about the nature of learning and knowledge formation direct the ways in which authentic assessment is interpreted and used. Educational issues that arise from some implementations of authentic assessment, identified as camouflage, simulation and abstraction, are discussed. The need for authentic assessment to be contextualised through a coherent teaching, learning and assessment domain is stressed.*

Learning theories recognise that learning is dependent on complex interrelationships of cognitive, affective and socio-cultural factors (Resnick, 1989). Context also has significant effects on learning and performance (Anderson, Reder & Simon, 1996; Wiggins, 1993). Assessment theory and practice have been evolving to reflect these complexities, moving away from more narrowly-focussed psychological theories of measurement that have dominated education until recently (Gipps, 1994; Goldstein, 1989; Linn, 1990, 1995). 'New approaches to assessment' have been identified as 'one of the major issues of the decade' (CTGV, 1993, p. 65).

The changing focus of assessment has led to two major theoretical considerations. The first relates to conceptions of validity, with renewed emphasis on the appropriateness of assessment tasks as indicators of intended learning outcomes, and on the appropriateness of interpretation of assessment outcomes as indicators of learning (Messick, 1989, 1994). These conceptions of validity are more compatible with the 'new paradigm of assessment' (Gipps, 1994), emphasising interpretations of quality and judgements of standards (Maxwell, 1997), than with measurement-oriented or psychometric approaches based on true score theory.

The second theoretical consideration relates to the need for learning and assessment of learning to be contextualised and meaningful for students. The quest for contextuality and meaningfulness arises from general awareness that learning and performance depend on context and motivation (Wiggins, 1993). Motivational benefits are expected to accrue when students can perceive the relevance of learning and assessment activities, thereby enhancing learning outcomes. This theoretical consideration incorporates concerns for transfer of learning from one educational context to another, from formal education to personal life and the workplace, and from life and workplace to formal education (Perkins & Salomon, 1989; Salomon & Perkins, 1989).

One development which attempts to address these issues has been the authentic assessment movement. The concept of authentic assessment has been embraced enthusiastically by policy makers, curriculum developers and practitioners alike, and enshrined in the literature on curriculum and assessment as a desirable characteristic of education. Of course, it is as difficult to be against authentic assessment as to be against apple pie and motherhood. It is obviously a 'good thing'. The alternative presumably would be in-authentic assessment and nobody would want that. Authentic assessment is clearly 'the way to go'.

This enthusiasm across many educational sectors has led to many different interpretations of authentic assessment. While it is natural for terms to evolve into common use and acquire different meanings, this can often lead to confused theory and practice. It is timely to consider the different meanings attached to authentic assessment and how such meanings relate to theories of learning and its assessment. Clarification of these matters should lead to better practice.

Authenticity

The first formal use of the term 'authentic' in the context of learning and assessment appears to have been by Archbald and Newmann (1988). Their reference was to 'authentic achievement'. In a later expanded treatment of their ideas (Newmann & Archbald, 1992), they state their focus as being the:

... fundamental question of what general forms of achievement *ought* to be promoted and assessed. ... Our concern here is not mainly with the technical problem of designing assessments that measure more validly what schools try to teach. ... What counts for success in school is often considered trivial, meaningless, and contrived – by students and adults alike. ... Ultimately then, the quality and utility of assessment rest upon the extent to which the outcomes measured represent appropriate, meaningful, significant, and worthwhile forms

of human accomplishment. We synthesize these qualities into one idea: authenticity. (p. 71)

Newmann and Archbald (1992) identify authentic achievement as having several characteristics that emulate the 'kinds of mastery demonstrated by successful adults' (pp. 72-74):

- Production of knowledge instead of reproduction or response only to the produced work of others.
- Disciplined enquiry, dependent on:
 - a prior knowledge base (to be used to help to produce knowledge);
 - in-depth understanding (rather than superficial awareness);
 - integration – the production of knowledge requires the ability to 'organize, synthesize, and integrate information in new ways'.
- Value beyond assessment – 'aesthetic, utilitarian, or personal value'.

Additionally, Newmann and Archbald (1992, p. 75) indicate that a major goal for authentic achievement is 'to cultivate the kind of higher-order thinking and problem-solving capacities useful both to individuals and to the society. The mastery gained in school is likely to transfer more readily to life beyond school ...'.

That is, authentic achievement should involve constructive learning, disciplined enquiry, and higher-order thinking and problem-solving. It should also have a value dimension, of aesthetic development, personal development or usefulness in the wider world. The last of these implies transfer of learning.

It was not long before the term 'authentic' was attached to assessment rather than achievement. The first to refer to 'authentic assessment' appears to have been Wiggins (1989). Although this transferred epithet appears natural and unexceptional in the context where the concern is not only with appropriate learning but also with how it is to be recognised, that is, with its assessment, there are implications for the way in which the discourse proceeds. 'Assessment of authentic achievement' places an emphasis on the nature of the achievement, or the learning, whereas 'authentic assessment of achievement' places an emphasis on the manner of assessment and could leave the nature of the achievement itself unexamined. In the following discussion, authentic assessment is more often referred to than authentic achievement as it has become the common parlance. However, it will be argued that it is important to examine carefully the nature of the learning which is anticipated or desired and to tailor the forms of assessment to that learning. Authentic assessment is not possible without attention to authentic achievement. A separation of the two can lead to empty rhetoric and facile assessment.

The relationship between theories of learning and authenticity

In their theory of educational goals, Newmann and Archbald (1992) argue that assessment should focus on achievement of authentic learning outcomes. The focus adopted for assessment has a substantial impact on the instructional process and the realisation of intended learning outcomes through the channelling it encourages of teacher and student effort. That is, assessment signals to teachers and students what is important in learning. Vice versa, assessment tasks need to relate to and take account of actual teaching and learning processes. Relationships between learning goals, teaching activities, learning processes and assessment procedures can be depicted in the form of a tetrahedron as in Figure 1. In a such a system of inter-relationships, all four components are in dynamic tension or balance. That is, adjustment of one component requires sympathetic adjustment of the other three. Sympathetic adjustment implies alignment of the underlying rationale or theoretical

assumptions of each component. That is, systemic validity inheres in the consistency of the underlying rationales of the four components. In particular, in terms of authenticity, the underlying rationale or theoretical assumptions relating to the authenticity of the learning goals and the assessment procedures not only need to match each other but also those relating to the teaching processes and the nature of learning and achievement.

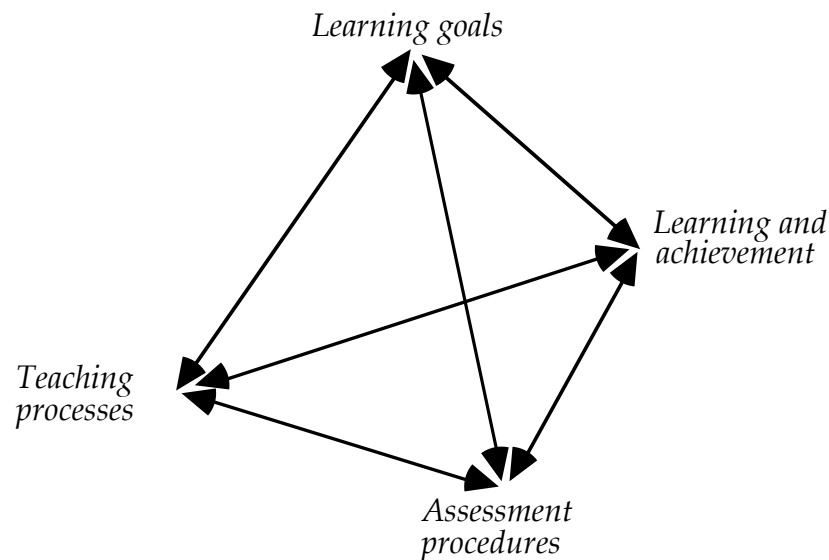


Figure 1: The teaching, learning, assessment domain

The authentic achievement outcomes indicated by Newmann and Archbald (1992) are drawn from a cognitive theory of learning and performance. The content of their discussion refers to cognitive processes and generic abilities such as higher-order thinking, embedded in a prior knowledge base. While Newmann and Archbald (1992) have extracted the essence of current cognitive learning theories, nuances in variations of these theories are important. Such different theoretical interpretations, particularly regarding the nature of knowledge and learning, lead to variations in the constructions of authenticity and the implementation of authentic assessment. Four major interpretations of authentic achievement and authentic assessment relate to:

- performance and performance assessment;
- situated learning and situated assessment;
- complexity of expertise and problem-based assessment; and
- competence and competence-based assessment.

Performance and performance assessment

A dominant construction of authenticity identifies authentic achievement and authentic assessment with performance assessment (see, for example, Torrance, 1995; Wiggins, 1993). Performance is the execution of some task or process which has to be assessed through actual demonstration, that is a productive activity (Wiggins, 1993). Performance also involves an emphasis on integration of knowledge and holistic applications. An holistic performance is one which requires attention to the whole task, not just separate pieces of it. For example, ability to write a report is assessed through actual performance in writing a report rather than through separate or generic skills such as writing a topic paragraph or correcting syntax or spelling. A task such as writing a report captures the notion of contextualised performance as well as holistic performance. Contextualisation here means making the task apparently 'real' rather than apparently artificial.

An important reason for focussing on holistic tasks is that performance on the whole task may differ from performance on component skills. For example, writing a report demands skills of integration which go beyond drafting of separate paragraphs. That is, the whole or Gestalt is not simply the sum of its parts. Transfer from performance on component skills to performance on the whole task should not be assumed.

This construction of authenticity assumes that assessment of performance involves only direct observation of the performance. However, assessment in such situations goes beyond observation of the performance to inference of generic knowledge, skills and processes (to write similar types of reports). It is not feasible to assess students' performances on a diverse range of activities and generalisations are therefore restricted and possibly dubious.

Situated learning and assessment

A second construction of authentic achievement and assessment emerges from theories of learning which claim that learning occurs best, or perhaps only, within context (Brown, Collins & Duguid, 1989; Campione & Brown, 1990; Perkins & Salomon, 1989). These theories posit that learning and performance are so *situated*, that is, developed within a context, that the actual schema of the domain knowledge and processes are different for different contexts. Therefore, a performance demonstrated in one context may not be indicative of capability for performance in another context. In this case, it is necessary to assess within the relevant learning context. For example, research has shown that adults can undertake in a supermarket context mathematical computations that they cannot manage in formal school settings.

Anderson, Reder and Simon (1996) discuss a continuum of theories of 'situatedness'. That is, theories related to situated learning range from those that posit possibilities of transfer of learning across contexts to those that argue complete contextual dependence. Complete contextual dependence would mean that assessment could only be authentic if it occurred within the specific context in which the statement about the performance standard is to be made. No generalisation would be possible.

Complexity of expertise and problem-based assessment

The third construction of authentic achievement and assessment recognises the complexity of expert performance which goes beyond technical facility. For example, an engineer designing a bridge has an open-ended problem with constraints of geography, material, financial and other resources. A problem-solving approach is needed but one which addresses the open-ended nature of the problem. The derivation of a solution, the bridge design, requires the interplay of information-gathering, collaboration, negotiation, selection of optimal characteristics, and underpinning theoretical knowledge. For performance of such complexity there is little comparison with traditional expectations of school instruction and assessment, where the emphasis has been on the production of predefined 'correct' answers to predefined closed 'problems'. Similar considerations apply to all kinds of expertise, for example, bricklaying, acting, tennis, hairdressing, musical performance and welding.

An example of a focus on complexity of expertise and problem solving is the work of the Cognition and Technology Group at Vanderbilt (CTGV) (1990, 1992a, 1992b, 1993). This group has developed instructional programs and assessment schema known as 'anchored instruction' which are directed at the development of expertise where 'instruction is anchored (situated) in ... problem-solving environments that teachers and students can explore' (CTGV, 1990, p. 2). Their work derives from situated learning theory but in addition emphasises emulation of expertise through problem solving. This problem-solving aspect of expertise reflects one of Newmann and Archbald's (1992) characteristics of authenticity, that is, the focus on higher-order thinking and problem solving. Newman and Archbald's other characteristics of

authenticity, that is, production rather than reproduction, disciplined enquiry (which includes a knowledge base, in-depth understanding and integration), value and meaningfulness are also present. This suggests that this approach is the most thorough-going realisation of the notion of authenticity so far.

It can be seen from this discussion that there are different learning theories underlying these three constructions of authentic achievement and assessment (summarised later in Table 1). However, these all include some attention to the development of higher-order cognitive processes rather than lower-order processes such as rote recall and production, and the importance of context in facilitating performance. These theories are consistent with the focus of Newmann and Archbald (1992) in their emphasis on complexity and context, even though other aspects of authenticity are sometimes absent.

Competence and competence-based assessment

Proponents of competence-based instruction and assessment have also embraced authentic assessment. In vocational education and training, *competence* is equated with satisfactory performance, particularly performance in activities and skills drawn from the work sector. Because of its direct ties to the 'real world', competence-based performance has become a fourth construction of authenticity, and *competency-based assessment* a construction of authentic assessment. The construction of authenticity in competence-based education is not drawn from theories of learning but from the theory that vocational education should be focused on direct application to the workplace and hence should reflect as closely as possible performance of the skills of the workplace.

Competence can be demonstrated at macro-levels – for example, in carpentry, the completion of a chair to a satisfactory standard – or micro-levels – satisfactory completion of mitre joints used in construction of a chair. While these levels appear to parallel the holistic versus component distinction of general education, such as writing a report versus sentence construction, in vocational contexts competent performance at either level might be appropriate. For example, in small businesses, employees may have multiple responsibilities whereas in large businesses employees may have very restricted activities.

Hence, demonstrated achievement of tasks at each level could be considered authentic outcomes in terms of 'adult performance' in 'real world' situations. However, the conditions of assessment in the vocational college for either level of task will not reflect the workplace directly. In the carpentry example, timbers, joints or furniture are sampled for assessment, not covered exhaustively, state of the art equipment and safety conditions are often present, and there is an absence of pressures related to financial considerations in business. The performance and its assessment are within a controlled environment. In this sense, authentic assessment is not possible until performance occurs in a genuine employment situation.

Table 1 summarises the discussion so far. It depicts the four major interpretations of authenticity and their realisation of the original conceptualisation of authenticity by Newmann and Archbald (1992). The table provides succinct description of the underlying learning theory of each interpretation, the intrinsic and latent components of Newmann and Archbald's conceptualisation, and significant additional features of each interpretation.

TABLE 1: Summary of four major interpretations of authenticity.

Interpretation	Underlying learning theory	*Intrinsic components	*Latent components	Additional features
Performance and performance assessment	Complex performance must be contextualised, holistic and integrated. The whole is more than the sum of the parts.	2 3c	1 3a, 3b (5a, 5b)	Realistic tasks. Communication.
Situated learning and situated assessment	Context of learning affects what is learned. Degrees of 'situatedness', with some arguing no generic transfer.	2 3a	3b, 3c 5a, 5b (6)	
Complexity of expertise and problem-based assessment	Draws on situated learning and expert performance with expectation of transfer. Use of complex scenarios to develop knowledge base and problem-solving skills.	2 3a, 3b, 3c 5a, 5b 6	1	
Competence and competence-based assessment	No apparent underlying learning theory.	4b	4c	'Real world'. Direct application to workplace.

* Components of authentic achievement as indicated by Newmann & Archbald (1992):

- 1 Appropriate, meaningful, significant and worthwhile forms of human accomplishment.
- 2 Production of knowledge, rather than reproduction.
- 3 Disciplined inquiry: a) prior knowledge base; b) in-depth understanding; c) integration.
- 4 Value beyond assessment: a) aesthetic; or b) utilitarian; or c) personal value.
- 5 Major goal to develop: a) higher-order thinking; b) problem-solving.
- 6 Expectation of transfer to life beyond school.

Performance and authenticity

Sometimes the symbiosis between definitions of performance assessment and authenticity can lead to each being seen to subsume the other, with tautologies developing. Some refer to authentic achievement and assessment as performance and performance assessment (Torrance, 1995, p. 1), while others such as Baker and O'Neil define performance assessment as incorporating aspects of authenticity. For example, Baker and O'Neil (1994, p. 15) describe 'performance-based assessment' as incorporating higher-order thinking and authenticity of *purpose* and elements of 'real world' performance:

... complex learning, higher order thinking, stimulation of a wide variety of active responses of students, tasks requiring multiple steps, and significant commitments of student time and effort. Performance-based assessment may also emphasize 'authenticity', that is, the task is intended to be inherently valuable to students, either immediately or because they can see its longer-term

connection to an important goal. It is also argued that performance-based assessment creates an opportunity for the integration of high quality subject matter learning into implicitly useful tasks. ... In practice, most efforts at performance assessment also require the student to communicate his or her understanding, of content, of process, and strategy, and of the results obtained. This communication component reinforces the 'real world' authentic aspects of tasks.

What is interesting to note about this quote is that 'complex learning', 'higher order thinking', 'active responses', 'multiple steps' and substantial effort are seen to characterise performance assessment and not authenticity, although these characteristics are closer to Newmann and Archbald's (1992) original definition. Instead, authenticity is identified in terms of utilitarian value and personal value to the student, which were part of the original Newmann and Archbald definition. Taken in total, Baker and O'Neil's definition of performance-based assessment is fairly close to Newmann and Archbald's definition of authenticity. Conversely, Newmann and Archbald's definition of authenticity is broader than Baker and O'Neil's, encompassing all of performance-based assessment, not just the value dimensions. However, Baker and O'Neil include communication and 'real world' as important aspects of performance tasks and appear to suggest that these are also components of authenticity.

Communication is also mentioned by Wiggins (1993) as an aspect of authenticity. He argues for assessments which 'better replicate authentic challenges and conditions instead of isolated drill exercises' (p. 210) and provides an example of 'how a performance for understanding differs from a test of knowledge' (p. 211). The example task, a document analysis, is for history students. The initial instructions and criteria for assessment are presented below.

You are a prosecutor or a defense attorney in a trial brought by a parent group seeking to forbid purchase by your high school of a U.S. history textbook excerpted below. (The book would be used as a *required supplement* to your current text, not in place of it.) You will present a ten-minute oral case, in pairs, to a jury, taking either side of the question, Is the book appropriate for school adoption and required reading? (supported by a written summary of your argument). You will be assessed on how well you support your claim about the accounts in the text, in response to the question. Are the accounts biased, inaccurate or merely different from our usual viewpoint? [History text excerpt follows]. (Wiggins, 1993, p. 212)

The task was structured for students by asking them to consider the following questions in their research and presentation.

1. What can be said to be the most likely political influences on the authors' point of view? What evidence is there of those influences? How do they affect the authors' choice of language? Does the language reflect bias or an acceptable (but different) point of view? Explain your reasoning.
2. Why does it make sense, given the authors' perspective, that they pay particular attention to (a) the Committee of Correspondence, (b) the contribution of women, and (c) the plight of "Indians" and "Negroes"? Are the facts accurate? Do they warrant that much attention in your view, or does such selective emphasis suggest a biased treatment? (How are these topics treated in the current text, and is the treatment there less biased or selective?)
3. You will be judged on the accuracy, aptness, and convincing qualities of your documentation, and the rhetorical effectiveness of *your* case. Be fair, but be an effective speaker and writer! A six-point scoring scale will be used for each

dimension to be assessed: persuasiveness of evidence, persuasiveness of argument, rhetorical effectiveness of speech, and support material. (Wiggins, 1993, p. 214)

This task involves both oral and written communication. It involves other aspects of performance tasks mentioned by Baker and O'Neil (1994), that is, complex learning, higher order thinking, active responses, multiple steps and substantial effort. In this sense, it also encompasses most of Newmann and Archbald's (1992) characteristics of authentic learning. However, there are some differences from the definitions of both Baker and O'Neil and Newmann and Archbald. It is worth considering this assessment task in some detail as it demonstrates many of the issues that have arisen in implementations of authentic assessment.

First, it is not clear that the task has personal and practical usefulness for the students. This is not a task of the kind that students would expect to undertake in their life outside school, either now or later. It is likely to be seen by students as being 'contrived'. Whether it is seen to be fun, challenging or meaningful by the students would depend on the student and a range of other contextual factors, and whether they are helped to understand the generalisable characteristics of such a task.

Second, the task extends the concept of 'real world' authenticity through the framing of historical and communication skills in terms of simulation of an adult professional role. Wiggins (1993) suggests that all assessment 'must always point toward and be "enabling" of adult performance' (p. 211). Even so, the emphasis in this task is not on accomplishing the adult role authentically, and a student who focussed on courtroom behaviour rather than on historical analysis would misinterpret the task requirements.

Third, the task is framed in such a way that students must accept the appropriateness and legitimacy of choosing a school supplementary textbook through court proceedings, and in one sense the notion of censorship. They must suspend any concerns they may have about such issues and focus on those which the teacher requires them to address. That is, they must clearly understand the demand characteristics and constraints of the task, some of which may be more implicit than explicit. Understanding what the teacher expects is always important.

Fourth, what is demanded of students in this task is not legal arguments but historical arguments. Students are directed to develop an argument about the 'appropriateness' of the textbook's historical interpretation and its 'accuracy' of facts. This requirement is the fundamental or *first-order expectation* of the task. The questions given by Wiggins in the task to guide the students' work direct the students' learning towards these historical issues. However, the task is embedded within a *second-order expectation*, performance within the simulated context of a courtroom, which overlays the student's historical understanding and analysis. The distinction between first-order expectations and second-order expectations is between expectations which relate to the underlying or core concepts, understandings and skills and expectations which relate to the specific context in which these are to be enacted or displayed. Here, the first-order expectation relates to historical knowledge, analysis and arguments while the second-order expectation relates to the emulation of legal forms, procedures and arguments.

While the learning focus appears to be on the first-order expectation, and the context is being provided mainly as a more interesting and engaging way for students to display their understanding of issues such as bias, historical representation and interpretation, the second-order expectation dominates the assessment criteria. That is, the assessment dimensions, or criteria, emphasise 'persuasiveness' and 'rhetorical effectiveness', which are characteristics of audience or reader reaction rather than characteristics of historical understanding per se. It appears ironic that although the historical analysis is directed at uncovering sophistry in historical argument, the focus

on second-order characteristics of the task could possibly encourage the students' own sophistry in argumentation.

These critical comments are not directed at undermining the use of complex, engaging and meaningful tasks which emphasise authentic learning. The movement of assessment in this direction is desirable. As noted earlier, Wiggins's example task has many desirable features of authentic learning and should encourage further experimentation in extending the range of assessment procedures. However, such tasks need to be carefully designed to ensure that student learning and teacher assessment are both focussed on the fundamental or first-order expectations of student performance. It is all too easy for both the student and the teacher to miss the point of the task. That is, it is all too easy for the second-order expectations to overwhelm the first-order – a case of the medium becoming the message.

A further issue that arises from the use of such complex performance assessments is the identification of instructional sequences which lead to successful performance. Overall, the instruction that justifies the assessment task and enables the students to complete it must be clearly delineated and matched to the learning expectations. In Wiggins's example, the four assessment criteria are 'persuasiveness of evidence, persuasiveness of argument, rhetorical effectiveness of speech, and support material' (Wiggins, 1993, p. 214). These are a mix of first-order and second-order expectations, mainly the latter, which would be better separated. However, the inclusion of second-order expectations in the assessment criteria implies that instruction would need to focus on developing students' skills of argument and presentation, both oral and written, as well as on selection and analysis of historical evidence. As implied in Figure 1, it would be inappropriate to hold students accountable for skills which they have been given no support for developing. Further, separation of the first-order and second-order expectations makes it more likely that the teacher will provide appropriate learning experiences in both. In Wiggins's example, that means distinguishing between a student's need to gain further insight into the nature of history and their need to develop rhetorical effectiveness.

In formative uses of assessment, this raises the question of whether the teacher should remediate separate components of performance or require the unsuccessful student to repeat the same, or a similar, holistic task until they succeed (as happens often with vocational competencies). The former may be necessary to address specific needs efficiently, partly because it is not possible to complete many complex tasks in an instructional program. However, development of capability in undertaking complex tasks requires practice in undertaking complex tasks, not just practice on component skills.

Camouflage

The constructions of authenticity previously discussed include to different extents some or all of the following: higher-order thinking, specific domain knowledge, and emulation of 'real world' activities, as well as communication of understanding, mastery, complex holistic activities, and emulation or facilitation of adult performance. Often, attempts to do 'authentic assessment' lead to superficial or inadequate implementation of these characteristics or to implementation in such a way as to simply place a gloss on existing assessment techniques. One such gloss we call *camouflage*.

Camouflage occurs when a traditional form of assessment is 'dressed up' to appear authentic, often by the introduction of 'real world' elements or tokenism. Wiggins's sample history task to evaluate students' understanding of history is to an extent camouflaged in performance characteristics of the law court, albeit characteristics that in this case could have 'enabling' value.

Camouflage occurs in attempts to implement authentic assessment, with varying degrees of sophistication and types of camouflage. The most flimsy are usually found in mathematics and problem-solving. A 1994 adult numeracy text, since revised, included examples such as the following multicultural camouflage:

Toula and Roula each buy a new book. Toula's book has 450 pages and Roula's book has 280 pages. If Toula reads 50 pages per day and Roula reads 40 pages per day, who will finish their book first?

Toula and Roula are recognisable names from an Australian television comedy show. However, their inclusion does not change the context of the problem or provide a degree of situatedness that facilitates a solution. Such camouflage may only serve to confuse. In fact, Cooper (1994), examining the boundary between common-sense, everyday knowledge and mathematical discourse in England's Key Stage testing, found that, in attempts to make tasks more 'authentic' or authentic in terms of pseudo-real life, the camouflage used often only served to make tasks more contrived and artificial.

Another example is found in Black (1991), derived from Schofield and others (1988). Students were given different practical tasks 'designed to be identical in respect of the structure of the tasks and of the demands that they entailed, except that the first (Sweets) was set as an everyday problem, expressed in everyday language and provided with kitchen equipment, whereas the second (Chemical) was set as a science problem with scientific terms and laboratory equipment' (Black, 1991, p. 18-19). The wording of the two tasks differed as follows :

SWEETS

When you eat certain sweets they dissolve in your mouth and fizz. They do the same thing in water.

This is what you have to find out:

What makes the difference to how long the sweets last for?

Is it how hot or cold the water is or how much the sweet is broken up?

CHEMICAL

When certain chemicals are placed in water they fizz and bubbles of gas are formed.

This is what you have to find out:

What makes the difference to how long the chemicals last for?

Is it how hot or cold the water is or how much the chemical is broken up?

In both cases students were instructed:

You can use any of the things in front of you. Choose whatever you need to answer the question. Make a clear record of your results so that someone else can understand what you have found out.

Separate samples of about 250 students undertook each task. Fifty-four per cent of students completed the chemical task satisfactorily but only 26 per cent of students completed the sweets task satisfactorily. The responses expected by the assessors required the conduct of a chemical experiment. It appears that the students undertaking the sweets task were distracted by the 'everyday context' and were not able to interpret it as a chemical experiment. The sentence 'They do the same thing in water' (that is, dissolve and fizz) did not provide the same orientation as 'When certain chemicals are placed in water they fizz and bubbles of gas are formed'. Many more students made inappropriate qualitative judgements with the sweets task. For many students clearly the way the language in which the task was presented camouflaged the real assessment expectations.

Rather than bridging the gap between school and life, camouflage introduces the additional task demand that students should realise that these tasks are not, in fact, 'real life' and that the examiners do have 'correct' and specific responses in mind. Hence, camouflage as a means of introducing authenticity to an assessment activity can be a form of deceit distracting students from the underlying expectations of the assessor. While such attempts to make assessment more 'authentic' appear to be directed at making assessment tasks more interesting for students, the orientation of students has to be considered. Students learn to 'do' school and school assessment tasks. When we change the rules, students have to re-learn the nature of what is required.

Simulation of the 'real world'

The desire to simulate the 'real world' in some implementations of authentic assessment also needs much more thoughtful consideration. Simulation attempts to offer 'life like' assessment activities. For many this has become the essence of authenticity although it did not figure in the original construction of authenticity by Newmann and Archbald (1992) who were concerned with learning that could transfer to a range of contexts. In common practice, simulation of the 'real world' has been added to, and in many cases has replaced, the original concept of authenticity.

An important characteristic of simulated tasks is that they are not actually 'real'. This is especially so if they are conducted in the classroom. But it is also the case if they are conducted in other contexts, such as work contexts, where the doing of the task does not have 'real' consequences, that is where the person being assessed is not fully accountable for the outcomes. For example, in the assessment of life-saving through a simulated rescue inadequate performance does not lead to anyone's death, and in the assessment of financial accounting through a simulated investment activity inadequate performance does not lead to anyone's bankruptcy. Another example is the assessment of teaching proficiency in a practice teaching situation, where the practice teacher is not ultimately responsible for the learning outcomes of the students and inadequacies can be covered by the classroom teacher. Furthermore, these activities are always undertaken under constraints which make the task 'artificial' not 'real'. That is, the task is 'framed' by the deliberate construction of limits to the task so that the task is not typically as open-ended or as unbounded as real world tasks where extra complications and unforeseen contingencies can arise.

Issues concerning assessment through simulation are discussed by Swanson, Norman and Linn (1995) in relation to the health professions where various approaches have been practised for many years. They discuss four different types of task simulation: patient monitoring scenarios; computer-based clinical simulations; oral examinations; and standardised patients. They identify eight 'lessons' for the educational community. While many of these lessons address scoring reliability and validity and differential outcomes for different contexts and types of assessment, two lessons are salient to the issue of simulation.

No matter how realistic a performance-based assessment is, it is still a simulation, and examinees do not behave in the same way they would in real life. (p.7)

Neither traditional testing nor performance-based assessment methods are a panacea. Selection of assessment methods should depend on the skills to be assessed, and, generally, use of a blend of methods is desirable. (p.11)

Swanson and others (1995) record that performance on simulations is found in general to be different from, and better than, performance in real life on the task being simulated. That is, performance on simulated tasks does not necessarily transfer to performance on real-life tasks.

In many educational settings, attempts to create simulations as a method of authentic assessment can prove impractical. For example, in one adult basic education course, the communication competence of 'making an enquiry based on information in a (newspaper) advertisement' was to be assessed. The teacher placed an assessor on a telephone away from the classroom. The student had to phone the assessor who pretended to be a rental agent. They then enacted the process of renting a flat chosen from a newspaper advertisement. The construction of authenticity was not complexity of task and higher-order cognition, but simulation of context through the newspaper advertisement and the distance of the assessor. The activity became time-consuming and unsuccessful because the assessor was unable to use visual cues to monitor the success of the student and to identify areas where difficulties were occurring. That is, where students have difficulties on such tasks, they can be extremely inefficient methods of assessment. The activity also overlooked the fact that in real life many people renting a property physically visit a rental agency rather than use the telephone.

This assessment task was initiated by a conscientious teacher trying to enact an authentic approach to assessment, where authenticity was interpreted as simulation of the 'real world'. Wolf (1995) has warned of the problems of such simulation in competence-based assessment. Research in the United Kingdom has demonstrated that overemphasis on performance simulation without differentiation of the cognitive demands of such tasks can have an impact on classroom instruction to the detriment of complex higher-order skills. That is, a focus on simulation can produce effects which are the reverse of those intended by Newmann and Archbald (1992).

Construct-centred authenticity and abstraction

Messick (1994, p. 17) suggests that there are two types of simulation in relation to authentic assessment: task-centred and construct-centred.

In the task-centered approach to authentic assessment, credibility depends on the simulation of as much real-world complexity as can be provided. ... The construct-centered approach ... (focuses) on selected constructs of knowledge and skill and the conditions of their realistic engagement in task performance. ... (A)spects of the test situation can be controlled or standardized. Such simulated tasks are authentic in that they replicate the challenges and standards of real-world performances and are representative of the ways in which knowledge and skills are used in real-world contexts, even though they do not simulate all of the complexity of real-world functioning.

This is a useful distinction, shifting the focus from replication of superficial characteristics of complete 'real world' situations to replication of their 'challenges and standards' and representation of the 'knowledge and skills' they require. As no simulation can replicate the real world performance, it seems more appropriate for teachers to use the construct-centred approach to authenticity advocated by Messick (1994), rather than task-centred authenticity. In this case teachers have to identify the most salient characteristics of the learning that they wish to foster and assess and then ensure that these are appropriately encapsulated in the assessment activity. Such a stress on the links between instructional purpose and assessment is fundamental to good assessment practice and more important than attempting to capture the superficial characteristics of the outside world.

It is now possible to return to and to re-analyse the history assessment task of Wiggins (1993). It can now be seen that this assessment task incorporates both construct-centred and task-centred approaches to authenticity. The construct-centredness of the assessment is seen in its focus on the abstraction 'understanding of history', that is, the concept that history is a representation of points of view and that different arguments can lead to different interpretations of events and hence different 'histories'. This

abstraction, which requires higher-order cognitive skills to be attained, is the construct being assessed. This was referred to earlier as the first-order expectation.

The task-centredness of the assessment is the use of simulated courtroom procedures which shape the overt characteristics of the student's performance. These characteristics were referred to earlier as the second-order expectation. However, as we have since noted, all such task-centred simulations fail to replicate the real-world situation. For, example, a law court would have a prosecutor and defence lawyer, considerably more interactions, such as with witnesses, and considerably more unpredictability. While the second-order expectation in this case entered into the assessment criteria and appeared to dominate the assessment, thereby distracting from the first-order expectation, the construct-centred and task-centred aspects of the assessment could have been separated. The motivational value of the task-centred aspects could then be recognised without the need for them to figure in the assessment criteria.

There are some unresolved issues concerning Messick's (1994) proposition. Basically, he suggests that authentic assessment should focus on the complex skills underpinning performance in multiple contexts, a partial return to Newmann and Archbald's (1992) conception of authenticity. However, Newmann and Archbald's important emphasis on meaningfulness and purpose for the learner is omitted. The perspective of the learner needs to be included for both ethical and motivational reasons. It is worth noting that Wiggins's (1993) history assessment task incorporates this aspect of authenticity.

Further, Messick's construct-centred approach involves abstraction of the knowledge and skills needed for application in real-world contexts and the construction of assessment tasks which allow the knowledge and skills to be demonstrated in controlled and representative situations. In this formulation, if taken too far, it is not clear whether there would be any essential difference between construct-centred simulation and many current problem-solving assessment activities. Neither would appear to address the context-dependency of performance, though this is an issue for which there appears to be no satisfactory resolution at present.

Conclusion

In this paper, we have raised a number of educational issues relating to current implementations of authentic assessment. It can be seen from our analysis that the original intent of Newmann and Archbald (1992) concerning authentic achievement and authentic assessment has not been sustained in practice. On the positive side, the concept of authenticity has led to increased attention to validity in assessment and to imaginative innovation in assessment practice. However, many interpretations of authenticity verge on labelling, and exultation by association, rather than enactment of fundamental principles. Some interpretations of authenticity have the capacity to reduce rather than improve the quality of teaching and assessment. A return to fundamental principles is urgently needed.

Simplistic implementations of authentic assessment through camouflage, and complex but superficial implementations through simulation, miss the mark and do not improve educational practice. In addition, it is important that, as complex learning and assessment activities are devised in the name of authenticity, what we have termed second-order expectations should not replace what we have termed first-order expectations. Although different interpretations of authenticity probably cannot be avoided, it seems desirable to promote the original intentions of Newmann and Archbald (1992).

In Figure 1, the interdependencies between teaching, achievement, learning objectives and assessment were noted in what was termed the teaching, learning, assessment

domain. These interdependencies indicate the importance of attending to questions of educational values (what learning goals are desired), theories of learning (how learning is perceived to occur), theories of teaching (how learning can be facilitated) and theories of assessment (how learning can be recognised) and maintaining coherence and balance among their underlying rationales. The validity of assessment can be evaluated in terms of the extent to which the assessment relates to the ascribed educational values, learning theories and teaching theories as well as to the realisation of the desired assessment theory. This is an extension of the notion of ecological relevance (Block, 1988), that is, the attention to situation and purpose. This appears to provide a firmer basis for the concept of authentic assessment. Newman and Archbald (1992) focussed on desirable learning goals and these have found widespread acceptance. They did not, however, address the other components of the teaching, learning, assessment domain and there has been less agreement on how learning can be realised through these components. The contextualisation of authentic assessment within the teaching, learning and assessment domain is where future effort should be directed.

References

- ANDERSON, J. R., REDER L. M. & SIMON, H. A. (1996) Situated learning and education, *Educational Researcher*, 25 (4), pp. 5-11.
- ARCHBALD, D. A. & NEWMANN, F. M. (1988) *Assessing authentic academic achievement in the secondary school* (Reston, Virginia, National Association of Secondary School Principals).
- BAKER, E. L. & O'NEIL, H. F. (1994) Performance assessment and equity: a view from the USA, *Assessment in Education*, 1, pp. 11-26.
- BLACK, P. J. (1991) Social and educational imperatives for changing examinations in: A. J. M. LUIJTEN (Ed.) *Issues in public examinations: a selection of the Proceedings of the 1990 IAEA Conference* (Utrecht, Uitgeverij Lemma B.V.).
- BLOCK, R. A. (1988) Enhancing cognitive skills, in: R. A. FELLEENZ (Ed) *Cognition and the adult learner* (Bozeman, Montana, USA, Center for Adult Learning Research).
- BROWN, J. S., COLLINS, A. & DUGUID, P. (1989) Situated cognition and the culture of learning, *Educational Researcher*, 18 (1), pp. 32-41.
- CAMPIONE, J. C. & BROWN, A. L. (1990) Guided learning and transfer: Implications for approaches to assessment, in: N. FREDERIKSEN, R. GLASER, A. LESGOLD & M. G. SHAFTO (Eds) *Diagnostic monitoring of skill and knowledge acquisition* (Hillsdale, New Jersey, USA, Lawrence Erlbaum Associates).
- COGNITION & TECHNOLOGY GROUP AT VANDERBILT (CTGV) (1990) Anchored instruction and its relationship to situated cognition, *Educational Researcher*, 19 (6), pp. 2-10.
- COGNITION & TECHNOLOGY GROUP AT VANDERBILT (CTGV) (1992a) The Jasper experiment: an exploration of issues in learning and instructional design, in: M. HANNAFIN & S. HOOPER (Eds) *Educational Technology Research and Development*, 40(1), pp. 65-80.
- COGNITION & TECHNOLOGY GROUP AT VANDERBILT (CTGV) (1992b) The Jasper series as an example of anchored instruction: Theory, program description and assessment data, *Educational Psychologist*, 27, pp. 291-315.
- COGNITION & TECHNOLOGY GROUP AT VANDERBILT (CTGV) (1993) Anchored instruction and situated cognition revisited, *Educational Technology*, March, pp. 52-70.
- COOPER, B. (1994) Authentic testing in mathematics? The boundary between everyday and mathematical knowledge in national curriculum testing in English schools, *Assessment in Education*, 1, pp. 143-166.

- GIPPS, C. V. (1994) *Beyond testing: towards a theory of educational assessment* (London, Falmer Press).
- GOLDSTEIN, H. (1989). Psychometric test theory and educational assessment, in: H. SIMONS & J. ELLIOTT (Eds) *Rethinking appraisal and assessment* (Milton Keynes, Open University Press).
- LINN, R. L. (1990) Essentials of student assessment: from accountability to instructional aid, *Teachers College Record*, 91(3), pp. 422-436.
- LINN, R. L. (1995) *Assessment-based reform: challenges to educational measurement* (Princeton, NJ, ETS).
- MAXWELL, G. S. (1997) Teacher Judgement of Achievement Standards in Performance Assessments, in: *Proceedings of the Annual Conference of the International Association for Educational Assessment Montreal, Canada, 12-17 June, 1995* (Montreal, IAEA, Faculté des sciences de l'éducation, Université de Montréal).
- MESSICK, S. (1989) Validity, in: R. L. LINN (Ed) *Educational measurement* (3rd ed.) (New York, USA, Macmillan).
- MESSICK, S. (1994) The interplay of evidence and consequences in the validation of performance assessment, *Educational Researcher*, 23(2), pp. 13-23.
- NEWMANN, F. M. & ARCHBALD, D. A. (1992) The nature of authentic academic achievement, in: H. BERLAK, F. M. NEWMANN, E. ADAMS, D. A. ARCHBALD, T. BURGESS, J. RAVEN and T. A. ROMBERG *Toward a new science of educational testing and assessment* (Albany, New York, USA, SUNY Press).
- PERKINS, D. N. & SALOMON, G. (1989) Are cognitive skills context-bound?, *Educational Researcher*, 18(1), pp. 16-25.
- RESNICK, L. B. (1989) Introduction, in: L. B. RESNICK (Ed.) *Knowing, learning and instruction: essays in honor of Robert Glaser* (Hillsdale, NJ, Lawrence Erlbaum).
- SALOMON, G. & PERKINS, D. N. (1989) Rocky road to transfer: rethinking mechanisms of a neglected phenomenon, *Educational Psychologist*, 24 (2), pp. 113-142.
- SCHOFIELD, B., BLACK, P., BELL, J. F., JOHNSON, S., MURPHY, P., QUALTER, A. & RUSSELL, T. (1988) *Science at age 13: a review of APU findings 1980-1984* (London, HMSO).
- SWANSON, D. B., NORMAN, G. R & LINN, R. L. (1995) Performance-based assessment: Lessons from the health professions, *Educational Researcher*, 24 (5), pp. 5-11, 35.
- TORRANCE, H. (1995) Introduction, in: H. TORRANCE (Ed) *Evaluating authentic assessment*. (Buckingham, UK, Open University Press).
- WIGGINS, G. P. (1989) A true test: toward more authentic and equitable assessment, *Phi Delta Kappan*, 70, pp. 703-713.
- WIGGINS, G. P. (1993) *Assessing student performance* (San Francisco, USA, Jossey-Bass).
- WOLF, A. (1995) Authentic assessments in a competitive sector: Institutional prerequisites and cautionary tales, in: H. TORRANCE (Ed) *Evaluating authentic assessment*. (Buckingham, UK, Open University Press).