Journal of Educational Enquiry, Vol. 5, No. 2, 2004

The assessment of literacy: working the zone between 'system' and 'site' validity

Peter Freebody
School of Education, University of Queensland, Australia

Claire Wyatt-Smith
Faculty of Education, Griffith University, Australia

Abstract

In this paper we address questions regarding the assessment of literacy, exploring the idea that practical judgements made in specific communities of practice are the phenomena at the intersection of validity for (1) local sites and (2) administrative systems. We propose a template for considering the challenges facing systems, administrators and teachers as they work to validly assess the important aspects of literacy learning by students in schools. We identify and discuss some concerns and assumptions that are common across the two validity sites, and some that are distinctive.

Introduction

Governments and the media often express concern over the literacy levels of school students and of the adult community. This concern derives in part from a belief in the significance of literacy for general success in school and later training, for occupational opportunities and for the economic and cultural well-being of society. In recent times, literacy has been considered one of the cornerstones of the 'new' economic order. It will be necessary to retrain workers in this new order as the workplace becomes more information-dependent and -saturated, and as people change jobs more frequently than in former times. Thus, investment in literacy education is taken to be a new-economy investment in the general level of human capital at society's disposal.

This 'human-capital' model thus locates literacy as a set of capabilities whose levels in the community and the individual have direct consequences for the economy of

a nation-state (Arnove & Torres 1999; Gee, Hull & Lankshear 1997; Lo Bianco & Freebody 1997). Schools, along with other education sites, are thereby centre-staged as regions of economic activity, also enjoying their traditional status as forces for cultural cohesion and personal betterment.

One response of government and educational authorities to literacy concerns has been to intensify and broaden the assessment of literacy. In many English-speaking countries, state and federal governments have begun to put testing and/or reporting systems of various kinds in place to make systems, schools and teachers more accountable for visible improvement in the literacy levels of their students. Most of these moves have been accompanied by some tying of system- and school-level funds or other resources to aggregated performance on literacy assessments, most explicitly in recent legislative moves in the US (see National Reading Panel site, accessed 2004). In these ways, literacy education and assessment have become high-stakes administrative issues. But precisely what counts as adequate literacy levels and how they might best be gauged remains a matter of considerable debate.

Our aim in writing this paper was to address questions about the assessment of literacy. We expand on the point that practical judgements made in specific communities of practice are the phenomena at the intersection of two validity sites: validity for *local sites* and for *administrative systems*. Some educators have challenged the capability of systems, state or federal, to validly assess the important aspects of the literacy learning of students in primary schools; and some have queried the validity of comparisons across schools, regions and states. Across these sites, considerable variation may be evident in syllabus emphases, in pedagogical approaches, and in the demographic characteristics of the communities in which the schools operate (Luke, Freebody & Land 2001; Wyatt-Smith 2003). Here we introduce these issues through examining the notion of validity as it applies in different literacy education settings – as it is 'done' in different institutional places.

Factoring in accountability

School systems have faced increasingly explicit and more deeply penetrating levels of accountability. In this context, a central question is: how can the teaching, assessment and reporting of literacy reflect genuine literacy demands outside of the test site – how do they 'recast' those demands as instructional, assessable events; how 'thick' can the simulation of 'real' literacy practices be? This question has often been heard as a challenge to the technicism of the bureaucratic agenda; the challenge seems to come either from an appreciation of how literacy learning and performance occur, embedded in and as part of local settings, or from a more general humanistic investment in the importance of students' idiosyncrasies and 'individual needs, learning styles' and the like.

However, this leaves literacy educators facing a choice between two unsatisfactory options. They must report at levels of abstraction and generality that run the risk of systematically undermining one of the major supportive motivations of literacy education – the 'human capital' proposition. Or, they must attend to and respect particularity at the level of theory, pedagogy and data collection – at the potential expense of system-wide comparability, thus jeopardising the defensibility of allocating support resources.

We argue here that the first choice characterises the work of a central administrative body of work as unreasoned and problematic in an educational sense. It foregrounds the unknowable diversity and localisation of teaching and assessment, such that they become intractable in the face of policy interventions. This leaves literacy educators impervious to proactive development, and thus structurally conservative in the face of the claims of comparability and accountable allocations of support. The other choice, we suggest, leaves us to characterise a teacher's on-site work as similarly unreasoned and systemically irresponsible, foregrounding the regime of the standard.

Naming the choices

To preview the discussion, the first choice – a reliance on the readily testable commonalities thought to be implicated in all literacy learning and teaching – implies that literacy educators are playing some unwritten but fundamentally comparable role in a particular enactment of what counts as literacy; a role that can isolate and thus celebrate or blame The Individual Student doing The Task. Here, the local setting is taken to mimic the controlled laboratory as an arena for fair and valid individual displays of standardised success and failure. A classroom, so defined, has free-floating Tasks that Individual Students encounter in hermetically sealed work zones that may be sufficiently distant from out-of-school notions of 'productive learning' that they can be seen as test zones for key competences.

In this arena, every act is at once an act of learning and display. Thus, the ultimate transformation of the classroom into the high-stakes laboratory is achieved: literacy activities become symbolically embodied in decontextualised tasks; purposeful activities become symbolically embodied in invitations for display; and people become symbolically embodied in profiles or lists of competences. All of these transformations are made possible by the 'common-sense' abstraction from specific literacy practices to generic literate competences. The cycle of governance by literacy assessment is complete when these competences are 'sold back' to the presumed clients of literacy education as ways in which *they* can think and talk about their or their child's strengths and weaknesses (or those of their child's teacher) in contrast with their expectations and aspirations. This process installs the powerful mystery of the standard, and ignores the local validity needed to give the human capital model of literacy education a better-than-even-money chance.

The second choice – emphasising the comprehensive local-ness of literacy learning and teaching – implies an analogous overwriting of the everyday educational and physical mobility of the clientele. It assumes that central agencies should play a role in supporting what they cannot know, with no reasonable sense of what can or cannot be supported, or of where their support may lead a given student or school. It installs the powerful mystery of diversity and dismisses any appreciation of the broad cultural and economic commonalities in which all literacy learners live. By resolutely domesticating and locally fixing the perceived literacy demands placed on people, this second choice also gives the human capital model of literacy education little chance; either empirically, or as a driver of defensible policy discourse.

By different routes and in different ways, both of these options – privileging the standard or the local – give the human-capital notion of literacy education little to go on. In the first case, what can be abstracted and thus tested does not necessarily correlate with what allows a person or a community to be productively literate on any given occasion. Gee (2001, pp xviii–xix) addressed this point, stating that:

> … a person with a specific type of literacy, suited for participation in a specific social practice, is a different sort of person (one with different powers) than a person without that specific sort of literacy … taking our gaze away from reading and writing ‘in general’ and turning our gaze to specific sorts of ways with words, deeds, actions, interactions, values, feelings, symbols, and tools within specific social practices.

In the second choice, a radical-contingency view of literacy serves to ‘neighbourhood’ the problem from a policy perspective and from the potential benefits of having a policy in the first place. Effectively, it abrogates collective responsibility as embodied in systemic policy formation, and neutralises attempts to develop shareable and defensible repertoires of effective practice. Moreover, in its strong form, it negates the role of policy makers as resources, and their allocation of resources to enhance collective practice.

Garfinkel (1967) has commented on the tension between the two choices, highlighting the distinction between objective and indexical propositions, and the extent to which one can be substituted for the other. Garfinkel argues that indexical propositions gain their significance only from the local circumstances they are made in or apply to, whereas the sense of ‘objective’ generalisations may arise from understandings of local circumstances, but can unequivocally apply to a variety of settings with ‘standard’ circumstances. According to Garfinkel, the substitutability of objective for indexical propositions can be viewed as a touchstone that distinguishes the exact from the inexact sciences and, even more fundamentally, as a practice of reasoning that makes the exact sciences possible.

Garfinkel (1967) went on to characterise the social sciences as being divided, conceptually and methodologically, on the question of whether the local circumstances that give propositions their sense are what he termed a ‘nuisance’ (p 6) for the

researcher, or a necessary aspect of the accounts researchers need to report. According to Garfinkel,

> attempts to rid the practices of a science of these nuisances lends to each science its distinctive character of preoccupation and productivity with methodological issues … [with] indexical expressions, by reason of their prevalence and other properties, [presenting] immense, obstinate, and irremediable nuisances to the tasks of dealing rigorously with the phenomena of structure and relevance in theories of consistency proofs and computability. (p 6)

Accordingly, Garfinkel (1967, p 6) argued that 'in attempts to recover actual as compared with supposed common conduct', and '*wherever practical actions are topics of study'*, 'the distinction and substitutability of objective for indexical propositions is always accomplished only for all practical purposes' (italics in original).

In these terms, the practical task of validly generating reports of students' literacy performances – through, for example, an educational authority, always conducted within and for the particular structures and relevances of local circumstances – 'remains unrealizably programmatic' (Garfinkel 1967, p 5). The validity of such reporting is resolved only in the light of the question of its practical purposes. Samples can be made to stand for people; tests stand for situated, practical literacy activity; and test items to stand for ways of using and producing written materials. However, the key issue is not the unequivocal validity of these reports in some social isolation, but the relationship between what is lost in the process of such 'standing-for' and the rationale available to the public regarding the applications of this information (for a fuller discussion, see Wickert & Freebody 1994).

The interface of assessment activities and performance contexts

The notion that assessing literacy can lead to the solution of practical problems is what gives particular significance to potential mismatches between literacy assessment events and actual literacy events in and out of the classroom (Barton 2002). One apparently simple question arises: How to characterise these 'actual' literacy events?

An example of an approach that emphasises the local is found in the work of Heap (1987), who has drawn attention to the inadequacy of attempts to assess functional literacy in daily life among adults. In particular, he highlighted the difficulty of simulating real-life problems related to literacy in a formal testing situation. Heap found that daily problems relating to literacy are characterised by a movement back and forth between the text and the concrete demands of 'doing' the task, rather than by a single reliance on a self-evident textual problem. This led him to characterise the use of literacy in daily life as '*text-aided functioning*', in which the most rational means of solving a particular problem or pursuing a particular goal may not be to rely solely on the written word provided, but to shuffle back and forth between trial and error attempts to solve the

problem and selected guidance offered by the text itself, or in collaboration with students or workmates.

Heap (1991, pp 28–29), drawing on Vygotsky (1978), made the following observation about 'local rationalities' in reading lessons:

> What counts as reading, procedurally, is whatever parties to a setting are apparently justified in believing to be the case about what reading is, what the skills of reading are, and how well any of the interactants performed. An interactant learns what reading is, how it is done, and what counts as reading, criterially, by paying attention to what counts as reading, procedurally, in particular situations ... whatever the teacher permits to pass, uninterrupted and apparently unchallenged, as an adequate display of reading skill, counts, procedurally as adequate, until further notice.

The display of reading and writing 'skill' is not only produced procedurally in literacy lessons (formal and informal); but in particular versions of consciousness of a person aware of being in the process of learning to read (Freebody & Freiberg 2001) or, for instance, the individual who knows how to prefer certain types of fictional character over others (Baker & Freebody 1989; Wyatt-Smith 1997).

Heap's argument is that assessments typically ignore the everyday practical reasoning in which literacy practices are embedded in most sites except school. This is a particularly tantalising point considering the rhetorical weight given to what counts as a 'competency', and the generally unexamined assumption of its transferability to sites in and out of school. In the processes of assessment, each simulation entails the presentation of a naturalised moral order that supports the institutional order of the school or the training institution – partly by presenting an institutionalised set of literacy practices as if they were essential or basic, and by presenting the curriculum's movement through these practices as the description of a developmental progression.

Assessment activities in educational circumstances presuppose some idea of 1) what is going on here and now; and 2) what may be inferred to be the reflection of this activity *out of* the assessment context. Any assessment procedure can be read for the particular way in which its possible relevance may be recast in a context outside of testing and the school. At a simple level, this gloss may entail formatting or content. Or, in more complex ways, it may entail implied social circumstances, resources, or combinations of these and other features of the task or the displayed and assessable practical consciousness.

In contrast to the points made about classroom life, more formal assessment procedures conducted later in the chronology of school-style testing often reveal more about the institutionalisation of literate practices, precisely because they take for granted the 'natural' status of the recastings of self and task that have been outlined. The following example, designed for use in a test battery concerning levels of functional

literacy among adults, shows some of these recastings in action. The example is drawn from Kirsch and Jungeblut (1986) and is discussed in some detail by Heap (1987).

Example 4.7:

> *A manufacturing company provides its customers with the following instructions for returning appliances for service:*
>
> *When returning appliance for servicing, include a note telling as clearly and as specifically as possible what is wrong with the appliance.*
>
> *A repair person for the company receives four appliances with the following notes attached. Circle the letter in the note which best follows the instructions supplied by the company.*
>
> **A**
>
> The clock does not run correctly on this clock radio. I tried fixing it, but I couldn't.
>
> **B**
>
> The alarm on my clock radio doesn't go off at the time I set. It rings 15–30 minutes later.
>
> **C**
>
> My clock radio is not working. It stopped working right after I used it for five days.
>
> **D**
>
> This radio is broken. Please repair and return by Priority Post to the address on my slip.
>
> (Kirsch & Jungeblut 1986; cited in Heap 1987)

Just as teachers and students have to reconstruct certain practical reasoning capabilities in order to participate in classroom or assessment practices, literate learners in this test context need to share in or collude in particular systems of ideas or discourses that make reading and writing practices sensible – the cultural logic of this hypothetical event as it is recast into an anonymous test item.

As an example of what such an exercise might look like, Heap (1987) listed some aspects of what he termed 'mismatches' between the 'real' task ecology and the test item. He discussed these features in terms of the 'thickness' of the simulation, concluding that most test items constitute thin simulations of the literacy demands of civil, community, academic or vocational life, or of tasks in school curricula. Note that the mismatches that Heap identified, itemised below, pose various questions about the local rationalities of the test to the point where, in some cases, the 'test' has become merely a task referring only to its own ecology. First, Heap noted some differences that could be addressed to make a test more related to hypothetical real-world settings, as follows.

Task layout and the process of task completion: Where do the instructions come from? When might a reader, for example, recognise a note rather write one?

Familiarity of content: Has anybody who is likely to be a participant ever received notes of this kind? Moreover, has anybody ever had to nominate one out of a group of notes that 'best follows the instructions'? Do repairers or readers of any kind rate the relative efficacy of notes?

The consequences of the respondents' efforts: What might happen if the given answer was right or wrong?

The relative significance of the items: In the battery used by Kirsch and Jungeblut (1986), circling a selection of long-distance telephone charges is as significant, in terms of scoring, as reading the dosage level of a medicine for a child. Could some weightings be decided upon that reflect experiences in another context?

Heap (1987) noted that modifications could be made to minimise the mismatch between the test and the out-of-school task along these four lines. However, he also described differences necessary or intrinsic to the testing context that could not be modified to minimise the mismatch. For instance:

Preferred order of task performance: Some sequence of events needs to be directed, or the student can be left to complete the items in an order of their choice. In out-of-test contexts, the sequencing of the events is either determined to a large extent by the structure of the task, or is part of the problem of acting rationally in that task environment.

Time-boundedness: Some time restrictions need to be imposed on the task for testing purposes. In out-of-test contexts, a task may be left for days or weeks, until some other social or material supports become available. Again, trying to simulate this would make it difficult to directly compare performances between individuals, and redefine the task.

Frame pre-specification: In what respect does the provision of the instructions recast the nature of the task, apart from the actual content of the instructions?

Relevance of the observable field: In the test situation, the ideal respondent knows that everything visible in the task is relevant, and anything that cannot be observed in the test document need not be relevant. This is a crucial element in thinning out the simulation. Outside of the test environment, part of the competency relates to knowing which aspects of the resources available or possibly available might be relevant.

Availability of resources: Along with knowing what might be relevant, a feature of competency in an out-of-test task is the ability to seek out and effectively interact with any material and social resources that could be recruited to assist in completing the task. These availabilities – and importantly, the manner, degree and frequency of appropriate recruitment – are central features in the local rationality of the context. In tasks that are part of the test, additional material resources are not usually relevant, and social resources are never relevant.

This last point raises the issue of the necessarily embedded nature of managing tasks related to what counts as a 'competency' – in and out of school. In the processes of assessment, each simulation entails presenting a naturalised moral order that supports the institutional order of the school or the training institution; partly via the use of institutionalised sets of literacy practices as if they were 'essential' or 'basic', and via presentation of the curriculum's progress through these sets of practices as, at the same time, the description of a developmental progression.

Centering on context and judgement processes

The now commonplace view that literacy practices are given their shape partly by features of the immediate setting – that they are always socially situated and thus variable in important ways from site to site – has substantial consequences at this moment in the cultural and industrial history of most industrially developed societies. This is a moment when economic, social justice and multicultural agendas are making significant claims on the provision of education.

But the legitimacy of such claims is under considerable cultural and economic pressure. The currently energetic development of more detailed and site-based descriptions of literacy practices, in programs of ethnographic and linguistic research, comes at the very moment when bureaucratic agendas, in acknowledging the importance of literacy to many aspects of public and private life, are calling for more publicly defensible and accessible approaches to the teaching, assessment and reporting of literacy and the outcomes of that teaching effort. Crucially, such defensibility rests on the judgement processes of educators and the degree to which those processes simulate on-site, day-to-day evaluations of literacy and embody the goals of centrally developed

syllabuses and testing routines. (For recent research on judgement see Wyatt-Smith et al 2003; Wyatt-Smith & Castleton 2004.)

Accounts of preferred and possible outcomes of various literacy provisions are central to these differing notions of literacy. Several questions seem to distinguish the major positions that have been locked in perennial debate about the matter of literacy assessment among learners:

- Is there evidence of one sub-competence or set of sub-competences somehow underlying these practices?
- Or, are outcomes best regarded as terms referring to sets of specific activities, practices and dispositions that vary substantially in their makeup from site to site, community to community, and culture to culture?
- Further, what are the advantages and disadvantages of thinking in each of these ways and acting on either one of these premises?
- Specifically, what individual and social resources might be thought of as the 'driving engines' of literacy development, and what might be resources that 'come with' progress rather than being central to progress?

While definitions of literacy practices impose limits on the imaginations of educators and policy makers, they do not guarantee particular forms of educational or cultural intervention. In the ill-fated Experimental World Literacy Program (EWLP), UNESCO used the definition:

> A person is literate when he (sic) has acquired the essential knowledge and skills which enable him to engage in all those activities in which literacy is required for effective functioning in his group and community, and whose attainments in reading, writing and arithmetic make it possible for him to continue to use these skills towards his own and the community's development. (cited in Oxenham 1980, p 87)

This definition was enacted in ways that allowed the context of the program to have apparently counter-productive effects:

> In practice this apparently relativistic and functional definition of literacy has been largely associated with narrowly-defined programmes with work-related objectives, concerned with improvements in labour productivity ... Ideologically specific objectives had been disguised behind a supposedly neutral model of literacy as simply technical skills. (Baker & Street 1993, p 58)

So, definitions can be ignored or acted out differently in the implementation of programs apparently derived from them. A comment on the fate of UNESCO's definition and the EWLP is offered by the former Director of the Literacy Secretariat of UNESCO:

> While UNESCO had promoted what it called the ‘mass literacy campaign’ approach in its early years, it turned to a more targeted strategy, called ‘functional literacy’ programmes in the mid-1960s and early 1970s. When learners in these latter programmes discovered that the only ‘functionality’ involved was to make them better workers, the majority of these experiments failed. That UNESCO’s approach since the period has been to provide technical expertise and advice according to specific needs in specific contexts, is an indication of the world community’s pulse in recent years. No single solution can be applied across countries. Programmes and strategies must emanate from perceived needs within individuals and their communities. (Limage 1993, p 23)

Limage outlined several messages from the widespread failure of the UNESCO programs and experiments. Among them was that, in different cultures, literacy, as a complex set of social practices, is embedded in and with oral and other symbolic communication practices in different ways. Therefore, to affect the everyday lives of their clients, those planning instructional interventions need to seriously take these embeddings into account.

Further, an important theme in most current debates about literacy provision concerns, in one way or another, the degree and significance of differences – among learners’ goals and backgrounds; language use; teaching, learning and working sites; the perceived needs and competences of learners; and literacy education programs. These differences are overwhelmingly evident in the findings of ethnographic studies of literacy education. The task of presenting useful and empirically defensible generalisations about literacy education is set against the backdrop of these differences. These differences are always visible and consequential, but generally under-theorised.

The failure of the experiments mentioned above, among other developments in education, has led to an emphasis on targeting individual client’s perceived needs (eg Spener 1994; Stein 1995), and a strong belief that there is no simple or single solution. Over the last 15 years, in some circles these attitudes have led to an instructional and administrative eclecticism and a resolute ‘localism’, based on a naive belief that the client can and will articulate a comprehensive picture of his or her ‘literacy needs’. In turn, the counter-productivity of this has only recently been realised. While the location of needs in the individual has strong resonance with many theories of learning and the ideology of individual voluntarism, the focus has been shifting toward communities’ versions of what counts as literate practice, and the broader cultural and economic conditions that sustain those versions.

Generic, key and underpinning literacy

Some educational researchers and theorists have persistently sought ‘underlying’, ‘core’, ‘generic’ or ‘key’ capabilities in human performance. Many of the statistical techniques receiving heavy duty in educational measurement entail the reduction of multiple indicative performances to apparently underlying dimensions – factors or clusters – defined statistically in terms of varying degrees of association. Approaches based on Item Response Theory are clear and prevalent examples (eg OECD 2000).

We will group at least three possible and distinct meanings of the set of terms under the heading 'key'. The first relates to the level of abstraction: given a number of indicative performances that correlate, what overarching abstraction captures their conjoined essence – a collective noun? The second meaning relates to exemplary status: given a collection of correlated or otherwise associated indicators, from which 'best example' can infer the essence of the others – a 'heavy loader' in the Principal Component sense? The third meaning regards marker status: which 'family resemblance' out of a set of correlated indicators is the most inclusive or defining in the collective?

Most people charged with producing standards for literacy assessment aim their work at the first meaning: can an abstraction be produced that is descriptive and generative of a range of highly related literacy activities? The abstraction, it should be noted, is also meant to perform the work of the other two meanings, or at least make them possible. That is, the abstraction should be representative and permit the ready generation of good and less good examples. In these ways, the abstraction is intended to be at least potentially equally valid across sites it may be involved in. It should also be a fair and, importantly, salient description of the practices in their sites of origin. What the abstraction does not cover should lack interest for literacy educators or at least not be directly pertinent to their interests.

A central issue is: what is left out of an abstract statement about literacy may be the very features that, as they are locally enacted, make its apparent instructional implementation valid on the site and in terms of the policy's intentions. We may choose to question the possibility of producing abstract statements that can remain faithful to the everyday practices they claim to describe and stimulate.

Or, more significantly, but finally less damaging to the overall project, we may ask: warranted by what theory and set of principles are the terms of this abstraction a pertinent and, for the literacy educator, comprehensive account of the practice? That is, what systematic ideas about the workings of reading and writing practices in everyday life give sense to the generalities included in, and excluded from, the competence statement? This question asks for the systematic elaboration of what might seem commonsensical about reading and writing. The process of abstraction often relies on assumptions that have not been explicated, and the appeal for a principled status for the abstraction is made to apparent common sense. So, what is the problem with commonsense?

Our point here is not focused on the impossibility of the principled production of abstractions about literacy events, or on the 'subjectivity' of the person producing the abstraction. Rather, our point is that, however obviously commonsensical the statement, and however widely people with shared 'guild' or insider knowledge (Sadler 1985) collaborate in its production, there are nonetheless assumptions being made about (1) how this abstraction articulates with everyday literacy events; and (2) how this

abstraction articulates with the everyday problems of reporting, allocation and policy properly faced by central funding and advisory authorities. The need here, therefore, is for explicit theoretical accountability in these two articulations.

The more an abstract statement of competence is disguised as transparently commonsensical and untheorised, the more open it is to diverging interpretations and enactment in instructional sites. These processes have at least two unsatisfactory outcomes: (1) they more inadvertently mislead the policy maker; and (2) there is a more tenuous and haphazard relationship between the policy's support for targeted programs that aim to enhance literacy competence, and the intellectual, social and cultural outcomes of those programs. Commonsense approaches to the generation of generic, core or key competences thus run the highest risk of making the human-capital model of literacy education look misguided. Nothing seems to work. Further, the commonsense nature of the abstraction offers the policy maker and the literacy educator nowhere to go in seeking an explanation of why 'more literacy competence' has not, in at least a subset of cases, led to better learning, schools and societies.

It is important to note that the agendas of the teacher/assessor and the policy maker, while divergent in many ways, are in fact necessary complements to one another's proper enactments. That is, the proper development of profiles, policy, and curricular guidelines and support depends crucially on principled, highly specific and divergent operationalisations in the practical sense and, more theoretically, also gains sense by principled and locally diverse interpretations. If these diverse enactments were not presumed, then, as a counter-balancing act, central authorities would require standardised testing; an option currently exercised or at least supported in several countries around the world.

By the same token, the effective development and implementation of locally relevant teaching and learning programs depends on an understanding of the plausibility of that process and how well it can be supported. Thus, the consequences of the validity and defensibility of these definitions of what is validly going on, and how those definitions are acted out, are substantial for the work of many people – most obviously for the students learning literacy, but equally for the providers, teachers, and policy and funding administrators.

The two agendas of literacy assessment

Generalised statements about literacy among learners form the bridge between a number of groups, often physically and institutionally separated, that are concerned with literacy provision. As such, these statements must play, fairly and productively, in a number of distinct games. One important task for contemporary literacy educators and those concerned with supporting them, therefore, is to work on a program that can untangle the various functions of competence statements and their implementation.

Two of the different games these statements are played in show the different working assumptions of the people charged with using them. For instance, we think (in good faith) that the work a literacy educator in a school or workplace can do with a profile, curriculum guideline, or standards statement must assume that the site is not comparable in important ways with others, and that therefore the statement must be inflected with diversity and site-specificity. On the other hand, the work that a policy maker interested in appropriate and productive allocations of support for literacy education can do with a competence statement must assume - again properly, and in good faith - that, for purposes of this work, sites can be viewed in significant ways as comparable; and that statements can be made about relative efficacy, however qualified and blurred by considerations about diversity in clientele and conditions.

In part, this paper has led to the need to separate two agenda: the everyday activities of literacy teachers and their needs for and uses of competence-based developments; and the everyday activities of literacy policy makers and bureaucrats and the ways in which they need to make principled uses of information generated by competence-based reports. Educators can benefit from explicit statements about the expectations of systems in the development of curricular modules and programs; the systematic diagnoses of specific literacy capabilities/resources that particular students should display and may need to develop; and the interpretable description of students for appropriate placement and direction into pathways of further learning. In turn, policy makers and bureaucrats can benefit from the use of broadly described capabilities in the support of effective, targeted programs that are defensible at a system and national level.

The two concepts of ‘validity’ in which these groups need to confidently operate differ substantially in an important sense. We might say that, in making competence statements operational, educators need to develop ‘site-validity’, and policy makers and bureaucrats need to develop ‘system-validity’. These two forms of validity have overlapping concerns about assessments reflecting locally relevant texts, tasks and contexts, and about the principled development of policy and allocation of support. However, their interests properly direct teachers and policy makers to do work that leads out from those common concerns in different directions. For each of these groups, what counts as productive statements about literacy capabilities/resources is and must be heard differently.

It seems that in many instances the issue is not to explicitly and deliberately want to privilege one or the other concern, but to hold a unitary view of validity that has not been explicated and is often a slippery conflation of the two concerns. If any sector within the literacy effort conflates these interests and forms of work, it thereby creates an oppositional arena in which the different sectors seem to contest the very nature of literacy practices and what constitutes effective literacy provision. The work of each sector of literacy provision depends critically on mutual appreciation of the forms of validity properly acted upon in each of the other sectors.

The specific context for that appreciation is the framework for reporting the collection of abstract statements about desirable literacy competences. Such a framework needs to be internally and externally coherent and plausibly comprehensive – that is, have plausible connections to the everyday literacy demands of a society, and to the everyday work of literacy teachers and central policy makers and bureaucrats. Below, we attempt to give guidelines for such a project, developing a number of recommendations about the possible content of reporting frameworks that satisfy the diverging demands of validity we have outlined above. These are drawn in part from Freebody, Cumming and Falk (1993) (in turn, partly derived from Freebody & Luke 1990).

First, a reporting framework should contain some statements about the *instructional contexts* in which literacy education takes place. These might include features of the physical and linguistic resources available to a program. Such statements inform a central funding agency and subsequent providers for a given client, and allow the funding body to be proactive in that it can expect literacy to be taken seriously as an instructional imperative across curriculums or training programs.

Second, a reporting framework should contain some statements about the *task domains* presented by the community, social, vocational or academic purposes addressed in the provision. While a central funding agency will not necessarily need to know or endorse particular theoretical or pedagogical approaches, it can nonetheless require that attention be given to the major domains or resources involved in developing proficiency in literacy and numeracy.

For instance, drawing upon Freebody and Luke (1990), Halliday (1992), and Lytle and Schultz (1990), a reporting framework should require statements concerning the range of resources called upon for participation in literate societies. Specifically, the framework should refer to the *procedural* resources (familiarity with coding and formatting conventions of a language and context); *textual* resources (familiarity with the powerful and prevalent common and specialised registers); and the *critical and cultural* resources called upon in common and important circumstances faced by the clientele (familiarity with the ways in which a culture conventionally uses literacy and numeracy practices to perform cultural and ideological functions).

An analysis of the functions of literacy practices within contemporary industrialised societies leads directly to the proposition that becoming a fully active member of such societies entails the development of complex and multifaceted literacy and numeracy practices. Further, this development involves learning about at least the basic symbolic codes and procedures, the textual activities and types, and the cultural and ideological dimensions conventional in any particular society. Theoretical positions, curricular materials or pedagogical practices that do not give serious attention to the three types of resources mentioned above do not reflect the kinds of work presented to literacy clients wanting to become fully active citizens in a contemporary literate society.

Third, a reporting framework should contain some statements about the *nature of progress* within the terms of these resources for literate practice. Literacy practices are important learning topics and resources in and out of formal training and schooling contexts in contemporary societies. In outlining domains of literate resources, perhaps the most theory-driven – and thus the most difficult feature for a project aimed at any portable framework – has been the production of statements about progress.

Some candidates for capturing a 'continuum' of progress have included 'developmental-psychology' based stages of literacy acquisition (Sulzby 1994), text-linguistic based models (Christie et al 1991), and cycles of 'dependence-independence-collaboration' in learning (Dixon-Krauss 1996). Each of these presents problems of cultural specificity/generalisability (including the question of cultural, social and individual differences); distinguishability in action (to what extent can operational demonstrations of 'progress' be empirically distinguished?); and comprehensiveness (do these descriptions represent a full view of the capabilities commonly taken to entail 'literate practice'?). They also pose problems of external logic – regarding their portability across clientele and educational contexts. However, these models of progress do not stand alone; they can be placed within the frames of instructional context and task domains. This framing takes the pressure off the more demanding forms of challenges to validity, and may well allow quite simple statements to be informative at the level required by central agencies or across systems.

Conclusion

In writing this paper, one of our overall aims was to derive the conclusion that opting for a hard version of either site or system validity – though this is attractive, possible and indeed visible in many contexts – can undermine the very reasons for trying to enhance literacy activities in the first place. As Garfinkel (2002, p 261) noted in his consideration of policy formation and conventional social sciences:

> The literatures of the social sciences movement are consistent, reasoned, clear, lucid, coherent, reproducible, teachable, correctable. IN any actual case, the claims are also wrong, unavoidably wrong …

We have pursued this argument by considering issues of the 'functioning' and 'relevance' of literacy educators and their clients and of literacy administrators and policy makers. The everyday recalcitrant problems faced by this latter group, we have suggested, are of just as much interest to the literacy educator and theorist as the problems faced by the teacher in the literacy classroom. Further, more explicit consideration of the everyday reasoning practices of both groups will show their common and properly distinctive interests, as well as their necessary interdependence.

Finally, we recognise that in this paper we have unproblematically divided the community of people concerned with literacy education into students, teachers and central policy makers and bureaucrats. Clearly there are many more interested and active

groups whose agendas we have not well represented in this paper – program coordinators, system supervisors, consultants, and several local, national and international professional associations with varied memberships and interests. While these interests cut across the issues we have discussed, they are nonetheless particular inflections of the problems of teaching and reporting, and deserve more specialised treatment than we could give them here.

We have tried to achieve a template for considering each of these levels of activity and the problems associated with them, and to clear some ground for a proper understanding of these problems that involves particular local procedures for reasoning about literacy, with some common concerns and assumptions across levels, and some distinctive ones. We suggest that conflating all of these concerns into debates that seem merely about the set of social practices we refer to as literacy is counter-productive to all concerned, especially those motivated to become more fully active members of contemporary literate societies.

References

Arnove R & Torres C (eds) (1999) *Comparative education: the dialectic between the global and the local.* Ladham: Roman & Littlefield.

Baker C & Freebody P (1989) *Children's first school books: introductions to the culture of literacy.* Oxford: Basic Blackwell.

Baker D & Street B (1993) The adult literacy and numeracy scales: theoretical issues. In P Freebody, J Cumming, I Falk & S Muspratt (eds) *A trialing and evaluation of the adult literacy and numeracy competency scales. Report to the Commonwealth Department of Employment, Education, and Training, Literacy/ESL Section,* vol 2, pp 54–68.

Barton D (2002) Literacy practices in local activities: a case study of an ecological approach to literacy. In J-P Hautecoeur (ed) *Ecological education in everyday life.* Toronto: University of Toronto Press, pp 137–149.

Christie F, Devlin B, Freebody P, Luke A, Martin J, Threadgold T & Watson C (1991) *Teaching English literacy: a project of national significance on the preservice preparation of teachers for teaching English literacy,* vol 1. Canberra: Department of Employment, Education and Training.

Dixon-Krauss L (1996) *Vygotsky in the classroom: mediated literacy instruction and assessment.* New York: Allyn & Bacon.

Freebody P & Freiberg J (2001) Re-discovering practical reading activities in schools and homes. *Journal of Research in Reading,* vol 24, pp 222–234.

Freebody P, Cumming J & Falk I (1993) *The trialing and evaluation of the adult literacy and numeracy scales.* Canberra: Department of Employment, Education and Training.

Freebody P & Luke A (1990) 'Literacies' programs: debates and demands in cultural context. *Prospect,* vol 5, no 3, pp 7–16.

Garfinkel H (1967) *Studies in ethnomethodology.* Englewood Cliffs: Prentice Hall.

Garfinkel H (2002) *Ethnomethodology's program: working out Durkheim's aphorism.* Lanham, Maryland: Rowman & Littlefield Publishers.

Gee J, Hull G & Lankshear C (1997) *The new work order: behind the language of the new capitalism.* Sydney: Allen & Unwin.

Gee J (2001) *New times and new literacies: themes for a changing world.* Paper presented at the 8th International Literacy and Education Research Network Conference on Learning, Spetses, Greece.

Halliday M (1992) Linguistic perspectives on literacy: a systemic functional approach. In F Christie (ed) *Literacy in social processes.* Darwin: NTU Press.

Heap J (1987) *Effective functioning in daily life: a critique of concepts and surveys of functional literacy.* Paper presented at the National Reading Conference, Arizona.

Heap J (1991) A situated perspective on what counts as reading. In CD Baker & A Luke (eds) *Towards a critical sociology of reading pedagogy.* Amsterdam/Philadelphia: John Benjamins, pp 103–140.

Kirsch I & Jungeblut A (1986) *Literacy: profiles of America's young adults.* Final report of the national assessment for educational progress. Princeton: ETS.

Limage L (1993) Literacy strategies: a view from the International Literacy Year Secretariat of UNESCO. In P Freebody & A Welch (eds) *Knowledge, culture and power: international perspectives on literacy as policy and practice.* London: Falmer Press, pp 23–34.

Lo Bianco J & Freebody P (1997) *Australian literacies: informing national policy on literacy education.* Canberra: Language Australia.

Lytle S & Schultz K (1990) Assessing literacy learning with adults: an ideological approach. In P Beach & S Hynds (eds) *Developing discourse practices in adolescence and adulthood.* Norwood: Ablex.

Luke A, Freebody P & Land R (2001) *Literate futures: review of literacy education.* Brisbane: Education Queensland.

National Reading Panel site (accessed 2004)

http://www.nationalreadingpanel.org/Publications/nochildleftbehind.htm

OECD (2000) *Measuring student knowledge and skills: the PISA 2000 assessment of reading, mathematical and scientific literacy.* Paris: Organisation for Economic Co-operation and Development.

Oxenham J (1980) *Literacy: writing, reading and social organisation.* London: Routledge and Kegan Paul.

Sadler DR (1985) The origins and functions of evaluative criteria. *Educational Theory,* vol 35, pp 285–297.

Spener D (ed) (1994) *Adult biliteracy in the United States.* Washington DC and McHenry: Center for Applied Linguistics and Delta Systems (available from Delta Systems at 800–323–8270).

Stein S (1995) *Equipped for the future: a customer-driven vision for adult literacy and lifelong learning.* Washington DC: National Institute for Literacy (ERIC Document Reproduction Service No. ED 384 792).

Sulzby E (1994) Children's emergent reading of favorite storybooks: a developmental study. In R Ruddell, M Ruddell & H Singer (eds) *Theoretical models and processes of reading* (4th ed), pp 244–280. Newark: International Reading Association. Reprinted from *Reading Research Quarterly,* vol 20, pp 458–81.

Vygotsky LS (1978) *Mind in society.* In M Cole, V John-Steiner, S Scribner & E Souberman (eds) Cambridge: Harvard University Press.

Wickert R & Freebody P (1994) *Accounting for context and competence in Australian adult education.* Presentation to the Organization for Economic Development, Philadelphia, PA, USA, April 1994.

Wyatt-Smith C (1997) Rethinking the achievement of the student writer. *Australian Journal of Language and Literacy,* vol 20, no, 3, pp 183–196.

Wyatt-Smith C (2003) Exploring the relationship between large-scale literacy testing programs and classroom-based assessment: a focus on teachers' accounts. In H Fehring (ed) *Literacy assessment: a collection of articles from the Australian Literacy Educators' Association,* pp 144–163.

Wyatt-Smith C, Castleton G, Freebody P & Cooksey R (2003) Teachers' qualitative judgements: a matter of context and salience, Part I. *Australian Journal of Language and Literacy,* vol 26, no 2, pp 11–32.

Wyatt-Smith C & Castleton G (2004) Factors affecting writing achievement: mapping teacher beliefs. *English in Education,* vol 38, no 1, pp 37–61.