Being Serious About Improving Student Outcomes: Bringing the Issues of Quality, Standards, and Teacher Judgement to Centre Stage

Claire Wyatt-Smith

Griffith University, Centre for Applied Language, Literacy, and Communication Studies

Currently, few would openly challenge the notion that assessment, teaching, and learning are fundamentally interrelated in good practice. Beyond this, however, there remains much debate about what should be assessed and how assessment should properly occur—in short, what assessment evidence should be most valued and used for reporting purposes. This presentation starts from two propositions: first, that large-scale standardised testing, of itself, does not lead to improvement; and second, that if we are serious about improving student outcomes, it is time to reassert the centrality of teacher judgement, explicitly defined standards, and literacy and numeracy as cross-curricular priorities. Drawing on a current study of teacher capacity building in assessment, participants will be invited to consider characteristics of "assessment for learning" including the matching of curriculum intent to both learning and assessment opportunities. Of special interest will be teachers' own accounts of how they have developed and worked with stated assessment criteria and standards in their classrooms, both to improve student learning and to judge the quality of student work.

Introduction

In the last two decades, among the numerous sociopolitical issues facing education in general and classroom teachers in particular, assessment is arguably the one most fraught with problems. Despite well over a century of research into measures aimed at devising and implementing testing procedures and scoring rubrics, the assessment of student achievement remains problematic, with increasing recognition of the limitations of current measurement theory and practice (Delandshere, 2002). Further, a combination of developments in the last few years has meant that recognition of the complex nature of how to assess with meaning has intensified. The first development that occurred throughout the 1990s was a focusing of research effort on expectations of schooling, as well as academic, workplace, and community practices. Assessment research brought to light, for example, how assessment is inherently a cultural (and therefore contextualised), value-laden practice, and that issues of socioeconomic diversity and gender, to name a few, can profoundly impact student outcomes. In terms of schooling, a key insight has been that children from diverse sociocultural and economic backgrounds bring to classrooms differing senses of the rules for using language and texts in particular settings

(Gee, 1990). A further development emerging in this period was recognition of how the traditions of teaching can and should "be reshaped to include and capitalize on the kinds of differences that children bring to classrooms" (Luke & Kale, 1997). Ironically, it was also during the 1990s that assessment outcomes, especially in literacy and numeracy, became an education policy priority, with governments in several countries including the United Kingdom, Canada, the United States and Australia demonstrating keen interest in large-scale standardized testing designed to generate quantitative data on what Freebody (2001) refers to as "the more obvious, quantifiable, generalisable and thus minimal features of individuals' management of written scripts" (p. 106). As McClay (2002) highlighted, this has resulted in teachers being "burdened with increasingly inflexible and inappropriate demands for decontextualised assessment of student work" (p. 53).

Taken together, these developments mean that teachers face competing demands in their classrooms. On the one hand, there are the imperatives to establish connections between in-school and out-of-school knowledges, ensuring that school activities are relevant to the demands of the larger world (Cumming & Wyatt Smith, 2001). On the other hand, as McClay (2002) highlighted, there is increasing downward pressure to demonstrate quality assurance and adopt narrow forms of assessment that stifle wide-ranging development. This situation has been exacerbated by the continued silence in assessment theory and research on the matter of how "critical pedagogy" "can be "done" in the course of managing the interactions between assessment and classroom learning" (Morgan & Wyatt-Smith, 2000, pp. 123–124). In short, what has not been established is how liberal pedagogic practices that encourage students to develop critical consciousness at the text face can articulate with assessment.

The issue in this paper is not to argue the strengths (and limitations) of attempts at critical-cultural approaches to pedagogy. Readers interested in these matters, especially as they relate to literacy, are advised to see Baynham and Prinsloo (2001), Barton (2001), and Street (1997). Instead, the aim is to explore what a framework for assessing student achievement might extend to if it were to aim for congruence with critical pedagogic practices (Gee, 1990; Street, 1993, 1997; Barton & Hamilton, 2000). The challenge therefore is to develop a framework that situates assessment alongside concepts of knowledge and achievement, as well as taking a sharp focus on the literacy demands of assessment, known to impact the quality of student outcomes.

What characterizes the assessment framework proposed later in the paper is the turn away from the longstanding distinction between assessment as being either objective or subjective, and away from the attachment of assessment criteria to the notion of scoring rubrics to regulate judgement and render it value-free. Instead, the turn is to a recasting of assessment to provide openings for multicultural, multivocal, diverse interpretations and ways of knowing and doing by making core to assessment teacher-student interactions and the issue of quality.

In what follows attention focuses first on the notion that teacher judgement is central to improvement and on the matter of what counts as evidence of quality, prior to proposing a framework for inquiring into assessment, outlined in terms of its constituent elements.

Focusing on teacher judgement as central to improvement

Assessment has become a major focal point for government, educational professionals, and the community, with support growing for the view that "an educational institution must increasingly be able to demonstrate to both itself and the world outside that it is fulfilling the aims that it has set for itself and the ones expected of it by society in general" (Broadfoot, 1987, p. 5). As suggested already, the last few decades in Australia has seen a marked proliferation of initiatives relating to assessment. Some have been primarily for diagnostic and monitoring purposes, some for measurement purposes, and some have been aimed at straddling both the diagnostic and measurement roles. Many of the large-scale census assessment initiatives have had and continue to have significant funding implications, with most involving the allocation of funding where students appear to be "at risk" of falling behind and in need of special intervention or extra assistance. The issue here is not with appraising the merits or otherwise of particular initiatives, but with making the point that today, from the earliest years of schooling, Australian students are faced with more "assessment moments" than at any other time in the history of educational assessment in this country.

A main challenge in providing quality assessment is to make clear the purposes of the various assessment programs that students participate in, and the coherence, among them—how they relate in terms of purpose, one to the other, across the years of schooling. A related matter is how and in what form vital assessment information travels with the student across the years, from state to state, from primary to secondary schooling, and at a local level, class to class. The need for clarity about purposes, coherence and the transfer of information applies equally to classroom assessments exclusively under the teacher's control, and to large-scale statewide, census testing programs (Wyatt-Smith, Cumming, & Elkins, 2005)

Studies have shown that there is a need to engage more widely in professional dialogue about the relationships between large-scale testing programs and classroom-based assessment and, more specifically, about the coherence between the assessment information that they provide (Wyatt-Smith, 2002). Little is currently known, for example, about how parents and teachers interpret and make sense of this information. However, what is known is that, if statewide testing programs are to have a genuine purpose of improving outcomes, then teachers need adequate support to ensure that they, and not the test, are the primary change agent. If there is agreement on this, then teacher judgement comes to center stage, valuing it in the understanding that it lies at the heart of good teaching and good assessment (Cumming & Maxwell, 2004; Maxwell, 2004). Typically, teachers welcome opportunities to participate in forums in which they share student work samples and discuss how they arrive at judgements. In such an active form of professional development, teachers have time to reflect on how they design and implement assessment opportunities, how they interpret the evidence that they collect, and how various sources and types of evidence are combined in the judgements that they make.

Judgement is a routine part of each teacher's work, and yet it is difficult to subject it to scrutiny, even by the individual teacher concerned, unless scaffolded opportunities are provided to do so (Phelps, 1989). Studies of teacher judgement have shown that individual teachers carry with them not only evaluative experience but, more specifically,

their own judgement policies that typically remain private, though they work to shape in powerful ways the processes by which judgements of quality are arrived at (Wyatt-Smith, Castleton, Freebody, & Cooksey, 2003). Moreover, operating in these policies can be valuation practices that are as much tied to recollected observations of in-class learning and behaviours, as to the qualities of the piece to be assessed. A way forward is to recognize that *teacher judgement* in conjunction with *clearly specified standards* and *moderation opportunities* are a linchpin of a robust assessment culture in schooling. Sustained professional conversations are vital aound matters including planning for assessment; how assessment activities are designed; how evidence is collected, interpreted, and recorded; what contexts are suitable for undertaking particular assessment activities; what standards are in place to assist teachers in assessing quality; and, finally, around what system support is provides for teacher decision-making in assessing student achievement. If this is accepted, then explicit provision for system-supported moderation or assessment consortia provide one means for bringing teachers together around actual student work samples. Then judgments can be de-privatised, the issue of quality can come to center stage, and judgements involving a process of matching work samples to stated standards can be made defensible in ways not otherwise possible.

The reality is that while many teachers have initiated their own professional conversations around assessment practice, both within their school and at district level, it is also fair to say that many teachers experience a sense of isolation as they go about their work as assessors, having no sustained opportunities for such sharing. A related observation is that the provision and proliferation of outcomes, in themselves, do not secure reliable judgements in which teachers and the community can have confidence. There is a clear and pressing need for supporting teacher dialogue around the issues of judgement, including standard setting, and how to make available for students useful information about expectations of quality.

Posing the question: What will we count as evidence of desired learning for the citizen of this century?

Currently, few would openly challenge the notion that assessment, teaching, and learning are fundamentally interrelated in good practice. Beyond this, however, there remains much debate about *what should be assessed* and *how assessment should properly occur*—in short, what assessment evidence should be most valued. Traditionally, examinations have involved the use of pencil and paper, timed tasks, and the student working, of course, alone and unaided by the teacher or others. It is fair to say that in some schools, and especially in the high stakes assessment years of secondary schooling, there exists strong confidence in timed, examination conditions as providing an assurance that the assessment evidence generated under such conditions represents "the truth" about student achievement — evidence that has not been tampered with or interfered with by the contributions of others. In the worlds outside of schooling, however, while individual performance under regulated conditions is no doubt valued in some workplace practices, also valued is individual perseverance over time—"stickability"—and how individuals collaborate to solve problems and collectively generate innovative approaches to achieve collective goals.

As we move into the 21st century, assessment policy and practice in schooling is being challenged to review the nature of the knowledges and skills being assessed. Also opening for review is the optimum range of contexts and conditions for collecting assessment information about how students work with and reconstitute knowledges. While students working solo, and therefore unaided by the teacher, may yield information about student achievement in this context, it does not and cannot provide useful information about how students work collaboratively on a shared enterprise to achieve group goals either at a point in time or over time.

The two related interests—knowledges and skills, as well as contexts and conditions—raise a suite of issues around how curricular knowledges are conceptualized and how different conceptualizations lead to quite different conceptions of achievement, as well as different assessment possibilities for students to demonstrate what they know and can do. Teachers in some school sites are already providing a diverse range of opportunities and conditions, including a balance between those assessments requiring solo demonstrations, and those involving pair or small group effort, in some cases, sustained over an extended period of time. Beyond this, the challenge is to extend assessment possibilities to take account of diverse contexts, actual and virtual, afforded by emerging information and communication technologies. Assessments that provide opportunities for students to work in multiple modes and channels of communication are not only desirable, but also necessary, with the expertise of the teacher and ICT support being critical in framing such opportunities.

There is a strong body of published assessment research emphasising that a hallmark of quality assessment for formative or improvement purposes is explicit provision of opportunities for students to develop evaluative experience and expertise (Johnson 2003; Sadler, 1987). Central to this position is the understanding that teachers carry with them latent or in-the-head knowledge of assessment expectations and when students are given access to such knowledge, they can lessen their dependence on the teacher as the sole arbiter of quality and provider of feedback. Consistent with this is the understanding that when students are equipped with explicit knowledge about the assessment criteria and standards against which performances are to be judged—as well as knowledge of how to use them for improvement purposes—students can become active participants in (rather than objects of) the assessment process.

The case for pedagogy to focus on assessment criteria and standards is widely accepted in the published literature, with Sadler's (1987) theorising of formative assessment being recognised as a seminal work in the field. Given developments in digital technologies, it is timely to revisit the case, especially in terms of how it presents the relationship between teacher and student. One underpinning assumption is that the teacher is the expert or connoisseur who knows not only how to recognise quality but is also capable of informing students about the features or characteristics of quality performance. The student is recognised as the novice who is to be inducted into the guild or insider specialist knowledge about quality and standards that the teacher has to offer. In effect, it is the teacher's role as insider of the guild to induct the student into knowledge about what constitutes quality, and in so doing, make the student an insider of the guild. In relation to new and emerging technologies, however, it can no longer be

assumed that it is the teacher who is the expert, with insider knowledge. It is more likely the case that the roles are reversed: the teacher as the outsider—the student as insider. For example, some students can readily take on the role of expert with digital technologies, with the teacher's knowledge about and hands-on experience of using these being considerably less than her students'. This observation opens the possibility for rethinking authority relations around assessment—for reconsidering how student-teacher relations may be reconfigured where expertise and a capacity to recognise quality do not necessarily lie with the teacher.

In these reconfigured relations, the teacher's claim to expertise may be tied primarily to how they promote both quality learning and the qualities of learners so that learning will increasingly be about creating a kind of person, with kinds of dispositions and orientations to the world and to ways of working with and reconstituting knowledge as problem-solvers and collaborators.

Proposing a framework for enacting assessment as inquiry

Consistent with the above call for new ways of assessing is Delandshere's (2002) notion of assessment as inquiry and her observation of how years of educational assessment research have presented arguments for new forms of assessment. Beyond this, Delandshere highlights how "the call for change in assessment follows an almost unanimous recognition of the limitations of current measurement theory and practice" (p. 1461). In engaging with the issue of assessment evidence to be evaluated (and therefore valued), I propose a four-part framework for enacting assessment as inquiry. The framework is necessarily a construct on my part and deliberately situates standards and teacher judgement in relation to four elements, shown below in Figure 1. These elements are taken as fundamentally interrelated to one another and also to how standards and judgement come to be enacted. The proposed framework, together with the discussion that follows, serve to put forward some initial thoughts about ways to explore the nature and form of assessment practices and will be further developed and elaborated in future writing.

Essentially, the proposition put forward is that, when assessment is understood as inquiry, the practices and processes of assessing—social and cultural acts of doing assessment in actual contexts—can be considered in relation to four main elements:

(a) conceptions of knowledge/s;

(b) assumptions about the relationships betweens assessment, learning, and teaching;

(c) judgement practices, especially as these relate to requirements of assessment tasks and expectations of quality performance (Sadler, 1989), and

(d) "curriculum literacies" (Cumming, Wyatt-Smith, Ryan, & Doig, 1998; Wyatt-Smith & Cumming, 2001), the term referring to the discipline-specific literacy demands that students meet in completing such assessment tasks, that typically remain implicit in teaching, learning, and assessment practices.

Each of these elements can be thought of as a lens that enables particular characteristics of enacted assessment to come to the fore. Collectively, the set of four

lenses work to reveal what is at play in how student achievement is evaluated and therefore valued. In this way, the framework has clear implications for identifying and examining the practices used to establish how quality is judged and reported. It is relevant to large-scale assessment programs installed by systems, as well as assessment that teachers undertake in classrooms to determine progress and to judge achievement. The focus is on identifying and examining the suite of conceptions, values, and assumptions at play in decisions about ways of doing assessment.

The proposed framework is prompted by the lack of a general theoretical position that connects assessment to meaning making (Delandshere, 2002), including concepts of knowledge, learning, and language. Each element of the framework is briefly discussed below and will be illustrated using authentic materials in the presentation.

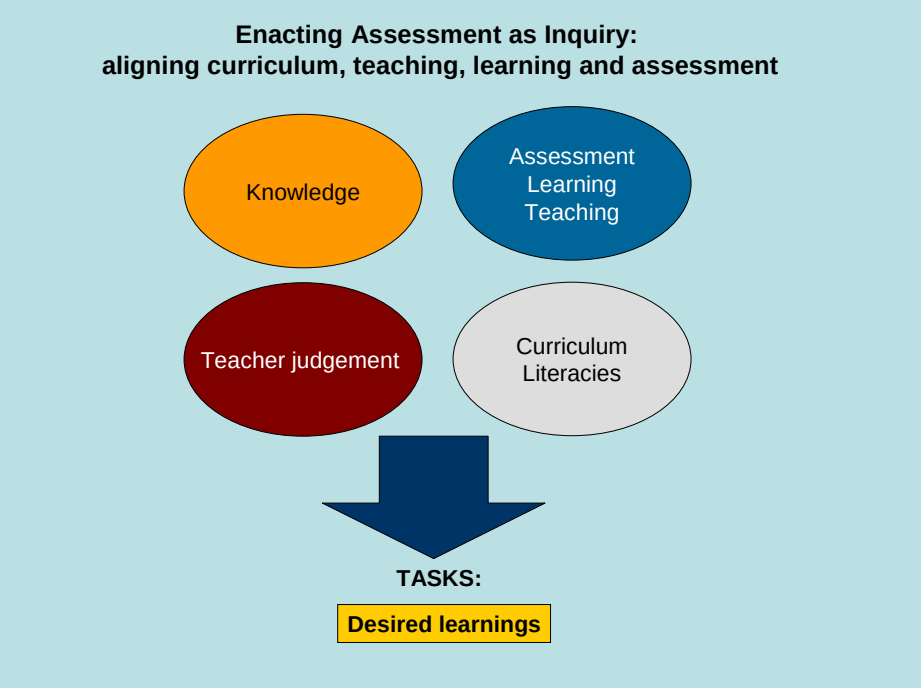

Figure 1.

Inquiring into assessment and enacting assessment as inquiry.

Lens 1: Conceptions of knowing and learning to be assessed

This lens brings to the fore conceptions of knowledge, and the assumptions about knowledge and learning that always and inevitably underpin acts of assessment. Despite the influence of such undergirding conceptions, their operation in and influence over

what comes to count as assessment evidence are rarely acknowledged. More than a decade ago, Gill (1993) made this observation, claiming that "Among the many and various articles and books on the quality and direction of American education, one searches in vain for an in-depth discussion of how knowing takes place, of who knowers are, and of what can be known" (p. 1). Drawing on this observation, Delandshere (2002, p. 1462) made the strong statement that

> Until we come to grips with, or at least frame the issue of, knowledge and knowing in ways that can guide education practices (including assessment), the enterprise of education runs the risk of being fruitless and counterproductive. In its current state, assessment appears to be a process of collecting data about phenomena or constructs that we have not adequately defined, to answer questions that we have not articulated, and on the basis of which we draw inferences about the quality of the education system.

Essentially, Delandshere's argument is that there is some urgency in reconnecting assessment and, more generally, educational practices to theoretical considerations as a means of clarifying assumptions made about what counts as valued knowledge, and therefore what should be provided for students in the name of quality teaching and learning. These two related matters raise a suite of issues around how knowledge, and more specifically curricular knowledges, are conceptualised and how different conceptualisations lead to quite different assessment possibilities for students to demonstrate what they know and can do.

Lens 2: Linking assessment, learning and teaching

In the last two decades, studies of assessment have shown increasing interest in how classroom assessment can be used to improve the learning experiences and outcomes of students. More specifically, the emphasis in educational assessment reform has increasingly been on meaningful, contextualised, and purposeful activity that focuses on demonstrations of what students know and can achieve, rather on students' shortfalls in knowledge and failure to achieve (Cumming & Maxwell, 1999; Gipps, 1994). Essentially, assessment has been re-framed in relation to its role in a learning culture (Shepard, 2000). As reviews of assessment and learning make clear (Black & Wiliam, 1998), the link between improved classroom assessment and the improvement of learning has been the subject of study by researchers from a variety of theoretical positions on teaching and learning.

However, in Graham Nuthall's (2004) critical analysis of why research has failed to bridge the theory-practice gap, he called for new research into learning, as it actually occurs in classrooms. He critiqued the predominant types of research in education and found that they were wanting in the area of real-time, direct, evidence-based data that "produce a practical understanding of how teachers' actions shape student learning" (p. 274). Indeed, it is significant tha,t while recent research trends have focussed strongly on notions of student (dis)engagement and on the multimodal nature of much classroom learning, research methods and forms of data collection in the field of classroom interaction lag behind in their applications of multimodality. Drawing on Nuthall's work, it is a fair observation that, traditionally, studies in educational practice including assessment have tended to rely on ethnographic techniques of data collection such as

observation notes, written reflections, teacher and student interviews, and classroom artefacts. While the use of audio and video recordings has been common, few studies have provided multiple, real-time recordings of whole class interactions. This has meant that, while the case is strong for aligning assessment, curriculum, teaching and learning, there is limited real-time, direct evidence-based data of the type identified by Nuthall. Further, previous attempts at audio and video recording of classroom interactions have been limited in the amount and type of audio and video information able to be collected and synchronised (i.e. single or dual track only). For education researchers, teachers and those involved in curriculum policy, the aim would be to study how teachers and students actually shuttle to and fro, across pedagogy, curriculum, and assessment, the constructs of knowing and learning that they jointly enact in so doing, and the fit between this and how assessment occurs.

Nuthall's (2004) research and his recommendations hold salience for researchers interested in the teacher-student-learning-assessment nexus. Of particular salience to a serious approach to inquiring into assessment practices are the recommendations calling for

(a) Independent in-depth assessment of what students learn.

(b) Complete, continuous data on individual student experience.

(c) Complete, continuous data on classroom activities.

(d) Analysis based on the continuous connections among classroom activities, student experiences, and learning processes,

(e) Avoiding the aggregation of data.

Lens 3: Teacher judgement linked to standards

Central to an inquiry approach to assessment is the principle that both the teacher and students both actively gather information about and reflect on learning and performance over time. Beyond this, teacher judgement is taken to be nested within a range of decision-making relating to curriculum frameworks, assessment practices, the school-community interface, and individual student learning needs and resources, both human and material. Generally speaking, there is support for this position both in the field of educational assessment research and in practice. In taking a focus on formative assessment, (Black & Wiliam, 1998b) stated that "there is strong support for the view that standards can be productive in informing not only judgement, but also teaching and learning". Drawing on a TIMSS video study (Stigler & Hiebert, 1997), they present the cautionary note: "A focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve."

The stance taken in this paper is that teacher judgement is best understood as evidence-based and that standards play a useful function in informing, substantiating, and making judgements defensible. Sadler (1987) stated,

> The primary function of educational standards is to enable statements about a student's quality of performance or degree of achievement to be made without reference to the

achievement of other students, which conceivably could be either all poor or all excellent. In addition, fixed standards enable long-term changes in a phenomenon to be detected.

Given the increasing education policy priority surrounding system access to "transparent" assessment information, there is no doubt that evidence-based judgements of achievement, measured against standards—and evidence of how such judgements meet the requirements for validity and reliability—are critical to continuing community confidence in schooling education. More than this, however, the challenges facing teachers charged with working with stated standards is to situate them in their classroom practice and, in so doing, take account of their school-community context. When this occurs, educational standards can become linked and interpreted in relation to pedagogy, curriculum and other school policy materials and also to the mix of knowledges, including knowledge of literacy, discussed next. Further, in the hands of teachers, educational standards can be used to inform pedagogy, the design of assessment activities that students are asked to undertake, feedback to students, and also students' self-improvement efforts.

Currently, little is known about teacher judgement and how standards can work to both inform and regulate judgement. A recent Australian study (Wyatt-Smith et al., 2003) showed, however, that in the absence of such standards, teacher judgement drew on a number of indexes, including assumed or included actual knowledge of:

(a) the community context in which the school is located

(b) colleagues talk about judgement

(c) assessment criteria and standards

(d) student

(e) pedagogy

(f) parental expectations, and

(g) teacher experience

The writers reported that, essentially, the indexes worked in two ways. First, they were the means that the teachers used to capture both the logic of judgements and the social relationships that they shared with their students. Second, they worked to constitute the judgements as rational and defensible. In effect, the index set can be understood as comprising a dynamic network of available knowledges that the teachers were reported to combine variously to account for how judgement occurs on a particular occasion. While there is no claim made in the study that this account of judgement practice is generalisable, it is clear that the insertion of defined standards into this practice, and the opportunity to moderate judgements, could provide a vital means for enhancing consistency of judgement practices.

Lens 4: Curriculum literacies

This fourth lens draws on a new conceptualisation of the literacy-curriculum interface that emerged from a national study of the literacy demands of curriculum in senior schooling (Cumming et al., 1998; Wyatt-Smith & Cumming, 2003). A key finding of the study was how it was "no longer appropriate to talk about literacy across the curriculum

or even literacy and curriculum" (Cumming et al., 1998, p. 12). Instead, the researchers developed the term "curriculum literacies", where "curriculum" is deliberately used as a noun, rather than the adjectival "curricular", to demonstrate that this conjunction represents the interface between a specific curriculum and its literacies, rather than literacies related to curriculum in a generic sense, or a single literacy that can be spread homogeneously across the curriculum. By way of example, the writers referred to Science literacies to describe the Science-literacy interface and, more specifically, perhaps Physics literacies, English literacies or Food Technology literacies (Cumming et al., 1998, p. 12).

Motivating this new conceptualisation of how literacy and curriculum relate in schooling is the finding that traditional definitions that construe literacy as primarily reading and writing did not match the observed literacy environment of a wide range of classrooms in which students were typically expected to coordinate multiple literacies simultaneously, drawing on listening, viewing, reading, writing, speaking and critical thinking (in order of apparent frequency) in complex and interrelated ways. A key point relating to this conceptualisation of curriculum literacies is that the focus on the literacies of curriculum areas does not lead to increased ambiguity. Instead, it is argued that it permits an increased pedagogic sharpening of what is meant by the term literacy in school subject areas. Specifically, the notion of "curriculum literacies" allows a shift away from

> the profligate use of the 'literacy' as synonymous with 'fluency' or a 'knowledgeable state' to the conceptualisation of literacies in terms of the integrating of reading, writing, listening, speaking, viewing, and critical thinking practices in recognisably-appropriate subject-specific ways (Cumming et al., 1998, p. 12).

Also of direct relevance to the discussion is the finding that the literacy demands of curriculum typically remained unstated or implicit in both pedagogy and assessment, including assessment for formative (diagnostic and improvement) and summative (reporting) purposes. Students were expected to manage these demands, in the main without explicit instruction. This was the case even though the demands themselves were shown to present powerful barriers to student learning and academic success. Building on this work, Wyatt-Smith and Cumming (2003) argued the need for exploring the coherence of literacy demands that students encounter in managing their learning in different contexts and the need to incorporate these demands explicitly in instruction and assessment. Their conclusion is that the reconceptualisation of curriculum literacies challenges current constructs of assessment and calls for the domains of assessment to be expanded to include both curriculum knowledge and epistemological domains that take account of diverse ways of working with and in semiotic systems. In a framework of assessment as inquiry, curriculum literacies are therefore central. It is this lens that focuses attention on the success (or failure) of systems as well as pedagogical and assessment practices to enable students to gain increasing control of this combination of curricular and literate knowledges and ability to use these productively.

Conclusion

This paper has opened up some of the complexities that can be considered when inquiring into educational assessment. It has proposed a framework through which to inquire into how assessment is enacted, both in classrooms and at system levels. At one level, the framework represents an attempt to see educational assessment in terms of its connectedness to issues of meaning: knowing, learning, teaching, and language. At another level, it is a provocation to reconsider the divergent assessment priorities and goals of the various education stakeholders in Australia and the pressure on some to follow short-term imperatives of appearing to be delivering improved results. Deep learning and improvement take time, however. They also involve new conversations around what is to be valued both in classroom-based and system assessment policies and practices. The challenge for the educational community is to be supportive of those assessment initiatives that focus on providing support for the long-term professional development necessary to effect change and deliver improved outcomes. As teachers know only too well, assessment procedures, of themselves, do not necessarily lead to improvement. Instead, teachers' professional knowledge and judgement practices are central, if we are serious about moves to improve student learning.

References

Barton, D. (2001). Directions for literacy research: Analysing language and social practices in a textually mediated world, *Language and Education*, *15*(2&3), 92–104.

Barton, D., & Hamilton, M. (2000). *Local literacies*. London: Routledge.

Baynham, M., & Prinsloo, M. (2001). New directions in literacy research: Policy, pedagogy, practice, *Language and Education*, *15*, (2&3), 83–91.

Black, P., & Wiliam, D. (1998a). *Inside the black box: Raising standards through classroom assessment.* Retrieved 21 November, 2005, from http://www.pdkintl.org/kappan/kbla9810.htm

Black, P., & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7–74.

Broadfoot, P. (1987). *Introducing profiling: A practical manual.* London: Macmillan Education.

Cumming, J. J., & Maxwell, G. (1999). Contextualising authentic assessment. *Assessment in Education: Principles, Policy & Practice, 6*(2), 177–194.

Cumming, J., & Maxwell, G. (2004). Profiles of educational assessment systems worldwide. *Assessment in Education: Principles, Policy and Practice, 11(*1), 89–108.

Cumming, J. J., & Wyatt-Smith, C.M. (2001). *Literacy and curriculum: Success in senior schooling.* Melbourne: ACER.

Cumming, J. J, Wyatt-Smith, C. M., Ryan, J., & Doig, S. M. (1998). *The literacy-curriculum interface: The literacy demands of the curriculum in post-compulsory schooling.* Final Report. Brisbane: Centre for Literacy Education Research, Griffith University, and Department of Employment, Education, Training and Youth Affairs.

Delandshere, G. (2002). Assessment as inquiry. *Teachers' College Record, 104*(7), 1461–1484.

Freebody, P. (2001). Theorising new literacies in and out of school. *Language and Education*, *15*(2&3), 105–116.

Gee, J. (1990). *Social Linguistics and Literacies* (2nd ed.). London: Falmer Press.

Gill, J. H. (1993). *Learning to learn: toward a philosophy of education*. Atlantic Highlands, NJ: Humanities Press.

Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London: Falmer.

Johnson, D. (2003). Activity theory, mediated action and literacy: Assessing how children make meaning in multiple modes. *Assessment in Education: Principles, Policy and Practice*, *10*(1), 103–129.

McClay, J. K. (2002). Hidden 'treasure': New genres, new media and the teaching of writing, *English in Education, 36*(1), 46–55.

Luke, A., & Kale, J. (1997). Learning through difference: Cultural practices in early childhood language socialization. In E. Gregory (Ed.) *One child many worlds: Early learning in multi-cultural communities*. London: Fulham Publishers.

Maxwell, G. (2004, March). Progressive assessment for learning and certification: Some lessons from school-based assessment in Queensland. Paper presented at the Third Conference of the Association of Commonwealth Examination and Assessment Boards *Redefining the Roles of Educational Assessment*, Nadi, Fiji.

Morgan, W., & Wyatt-Smith, C. M. (2000). Im/proper accountability: Towards a theory of critical literacy and assessment. *Assessment in Education: Principles, Policy and Practice, 17*(1), 123–142.

Nuthall, G. (1999). Introduction and background [Monograph]. *International Journal of Educational Research*, *31*(3), 141–256.

Nuthall, G. (2004). Relating classroom teaching to student learning: A critical analysis of why research has failed to bridge the theory-practice gap. *Harvard Educational Review, 74*(3), 273–307.

Phelps, L. (1989). Images of student writing: The deep structure of teacher response. In C. M. Anson (Ed.), *Writing and response: Theory, practice, and research* (pp. 37–67). Urbana, IL: NCTE.

Sadler, D. R. (1987). Specifying and promulgating achievement standards, *Oxford Review of Education, 13*, 191–209.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems, *Instructional Science*, *18*, 119–144.

Shepard, L. A. (2000, April). *The role of assessment in a learning culture*. Presidential address, American Educational Research Association, New Orleans.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An Overview of the TIMSS Video Study. *Phi Delta Kappan*, *September*, 19–20.

Street, B. (1997). The implications of the 'New Literacy Studies' for literacy education. *Language and Education*, *31*(3), 45–59.

Wyatt-Smith, C. (2002). Moving beyond 'English and literacy' to 'English literacies': Bringing together the study of texts and social practices. *English in Australia, 134,* 57–66.

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 26*(2)*,* 11–32.

Wyatt-Smith, C., Castleton, G., & Ryan, J. (2005, forthcoming). New research methodologies for researching new literacies. *International Journal of Learning.*

Wyatt-Smith, C. M., & Cumming, J. J. (2001). (Special Issue Editors). Examining the literacy-curriculum relationship. A Special Edition of *Linguistics and Education, 11*(3), 295–312.

Wyatt-Smith, C. M., & Cumming, J. J. (2003). Curriculum literacies: Expanding domains of assessment. *Assessment in Education: Principles, Policy and Practice, 10*(1)*,* 47–59.

Wyatt-Smith, C. M., Cumming, J., & Elkins, J. (2005). Opportunity to enhance learning: Redesigning assessment to sustain student engagement in the middle years. In D. Pendergast & N. Bahr (Eds.), *Middle years reform* (pp. 271–299*)*. Allen & Unwin.