

This article was downloaded by: [University of Waikato Library]

On: 18 November 2009

Access details: Access Details: [subscription number 907696412]

Publisher *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Language and Education

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t794297816>

### Literacy Assessment Practices: Moving from Standardised to Ecologically Valid Assessments in Secondary Schools

David Whitehead <sup>a</sup>

<sup>a</sup> School of Education, University of Waikato, New Zealand

**To cite this Article** Whitehead, David 'Literacy Assessment Practices: Moving from Standardised to Ecologically Valid Assessments in Secondary Schools', *Language and Education*, 21: 5, 434 – 452

**To link to this Article:** DOI: 10.2167/le801.0

**URL:** <http://dx.doi.org/10.2167/le801.0>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Literacy Assessment Practices: Moving from Standardised to Ecologically Valid Assessments in Secondary Schools<sup>1</sup>

David Whitehead

School of Education, University of Waikato, New Zealand

SSLI test protocol data revealed the dominance of 'central' literacy measures and 'local' subject-specific measures aligned to institutional requirements, curriculum and national examination content. These measures initiate secondary students into a pervasive culture of assessment that generally fails to support further learning; a culture antagonistic towards the use of assessment that reflect how expert teachers address subject-specific literacies. In a culture of content-focussed, high stakes assessment, the use of *ecologically valid* formative assessment that reveal what students can do with what they know, and that empower teachers to test like they teach, is marginalised. Consistent with Neisser's claim that some experimental measures may not reflect reality, the pedagogy and assessment protocols of many secondary schools fail to reflect the use of literacy and thinking tools, and so fail to reflect best evidence about teaching. Changes in school culture, teachers' pedagogical knowledge and the use of ecologically valid assessments are associated with shifts from transmission to co-construction approaches. Consistent with the work of David Corson the use of ecologically valid assessment that reflect the use of literacy and thinking tools is an inclusive, future-focussed literacy event, but the use of 'central' curriculum and institutional-linked measures is exclusive.

doi: 10.2167/le801.0

**Keywords:** assessment, ecologically valid, literacy and thinking tools

## Introduction

The blow of the psychometric cudgel wielded by secondary schools, which administer inappropriate central assessments of literacy, has become an all too common rite-of-passage for New Zealand students, as it has elsewhere (Smyth, this issue). These assessments initiate students into a curriculum culture characterised by content-focussed assessments linked to national standards, a regime that has a powerful influence on what teachers teach and how students learn. While these assessments might provide some measure of literacy achievement, or indicate students' ability to recall content, they rarely reflect the pedagogical tools used to facilitate learning.

An analysis of assessment practices in schools that participated in the *Secondary Schools' Literacy Initiative* (SSLI) from 2003 to 2005 confirmed the dominant role of the central forms of literacy achievement and local forms of content-focussed summative assessment in the majority of these schools (Whitehead

*et al.*, 2004). Central literacy assessment protocols tend to be standardised, quantitative, summative, and administered to meet the institutional needs for accountability. The use of central forms of literacy assessment has been challenged by those who regard them as decontextualised from the realities of the classroom and serving institutional needs rather than students' literacy and learning needs (Amrein & Berliner, 2002). Evidence obtained from our research evaluation suggests that the use of some of these measures was inappropriate because they were culturally biased, lacked construct validity, and because they were used inappropriately, for example, to stream or track classes.

Local, site-specific, summative assessments were, most commonly, in the form of teacher-made unit assessments, designed to measure the extent to which students had acquired transmitted knowledge. These assessments failed to reflect how the knowledge had been taught, or whether students could use what they knew. Although teachers controlled these assessments, they were rarely used formatively for learning in the majority of the secondary schools examined – a practice that is widely reflected elsewhere (Black & William, 1998; Clarke, 2001).

In contrast, a minority of SSLI schools used, in addition to appropriate central measures, local, site-specific, *ecologically valid* (Neisser, 1976) forms of assessment designed to measure both what was taught and the literacy and thinking tools used to teach the content. These assessments were consistent with a definition of knowledge as 'knowing', that is, as knowledge used to create new knowledge, rather than the knowledge-as-object metaphor that characterises most central forms of assessment (Gilbert, 2005). In these particular classrooms, the use of local, site-specific, ecologically valid assessments was accompanied by co-construction pedagogies, rather than the transmissionist teaching more characteristics of the sole use of central assessments.

The starting point for this paper is to acknowledge that measures designed to help teachers to test as they teach are at risk in secondary school cultures dominated by central institutionally focussed assessment protocols, justified on the basis of national, curriculum-linked examination systems, departmental autonomy and historical precedent. The point is also made that the potential combined value of central and local assessment data is yet to be realised in many secondary schools, particularly those schools that are in the early stages of development with respect to literacy across the curriculum (May, this issue). This latter point signals a key finding of the SSLI research around the progression of schools from reliance on inappropriate central measures to the adoption of diagnostic and site-specific, ecologically valid measures. The case study detailed later in this paper, drawn from the SSLI research evaluation, suggests the possibilities and benefits of using local, ecologically valid assessments that reflect teachers' use of literacy and thinking tools. These local forms of assessment are shown to enhance the engagement of teachers and students, while placing literacy at the heart of all subjects.

## Getting Real

The argument that central assessments of literacy achievement are not representative of socially contextualised literacy tasks that occur in classroom settings is not new (Freebody & Wyatt-Smith, 2004). The key argument here is that these

assessments fail to reflect the unique and actual literacy demands of subjects across the curriculum. Likewise, the argument that the administration of assessments under controlled conditions may not reflect real educational settings is well known (Brock-Utne, 1996; Neisser, 1976). It is questionable whether central forms of literacy assessment reflect how students acquire subject-specific literacies in secondary schools, and whether they predict student success on literacy tasks that are unique to each curriculum subject (Higginson *et al.*, 2000).

Associated with these arguments are more general issues of *construct validity*. With regard to central assessments of literacy achievement, this argument centres on the degree of alignment and transfer between the set of generic literacy skills selected by test designers, and the skills needed by students to cope with the demands of subject-specific texts (Barton, 2002). For example, some of the procedural skills central to the comprehension of narrative are different from those required to comprehend non-fiction, but these differences are unlikely to be reflected in the construction of central assessments of reading comprehension.

Together, these arguments focus on whether the methods, materials and settings embodied in central measures approximate real-life, pedagogical contexts. Implicit in this argument is that they may fail to meet the 'reality' criterion of the classroom.

## Locus of Control

In addition to these psychometric and contextual arguments, there are also political arguments around the appropriateness of central measures of literacy achievement. Again, these arguments focus on tensions between the local and central control of education (Garfinkel, 1967).

## Central control

The central assessments used in the SSLI schools examined as part of the research evaluation reflected institutionalised sets of literacy practice. Rather than possessing what Freebody and Wyatt-Smith (2004) describe as the 'site-validity' of local assessment measures, central measures possess 'system-validity' because they reflect the needs of policymakers and bureaucrats to profile 'literacy' defined in a generic sense. 'System-validity' is likewise reflected in content-focussed, subject-specific unit assessments. During the 2003–2005 period of evaluation, the SSLI researchers identified that in all but one of the 60 SSLI schools, some students, most often those deemed to be less literate, might be administered up to eight tests of literacy annually. The most common measures were summative tests of reading comprehension, vocabulary and listening, designed for annual administration in February.<sup>2</sup> Given this plethora of tests, there seems to be a case for a reduction in the number of central tests administered annually in secondary schools in New Zealand (Whitehead, 2006).

## Local measures

Most of the local assessment measures used in the SSLI schools, including measures of literacy, were linked to subject-specific content. The content validity of these local assessments was acquired by reference to subject-specific texts,

prescriptions and examinations. They were valid, in a psychometric sense, because of the close alignment between the content taught and the content tested. In contrast, the content validity of ecologically valid assessments reported in the following case study was acquired by reference to both the content delivered *and* the use of literacy and thinking tools used to teach and manipulate that content. These ecologically valid assessments were designed to measure the subject-specific declarative knowledge that the students acquired *alongwith* their procedural knowledge of literacy and thinking tools. These local forms of assessment reflected the real learning demands of the classroom – that is, both the subject-specific functional literacies, and disciplined-linked thinking events that occurred. Just as the sites where these assessments were administered were characterised by their own discourses and ways of constructing knowledge, the use of ecologically valid local measures were characterised by items that reflected that discourse and these events. The design of these assessments acknowledged that subject-specific literacy practices are given their shape within a curriculum and pedagogical context; that is, the assessment practices are socially and academically situated.

Although arguing for a place for the use of local, ecologically valid assessments, we need to acknowledge that, to satisfy the policy demands of centralised institutions and to provide schools with longitudinal data for comparative purposes, there is also a place for central forms of assessment. Whether these need to be constructed as ‘high-stakes’, as is currently the case in the USA, is another matter (Amrein & Berliner, 2002; Chaytor & Schmitter-Edgecombe, 2003). This paper argues for, and illustrates, how teachers can use local, ecologically valid measures along with these central assessment measures, and explains why some teachers in the SSLI project chose to add local, ecologically valid assessments to their traditional subject-specific assessment protocol. It also argues for an association between the emergence of local assessment measures, school culture, pedagogy and the confidence of teachers to use literacy and thinking tools.

### Use of Central Measures of Literacy Achievement

In an era dominated by central testing (Amrein & Berliner, 2002) and frenetic levels of government policy, consternation over matters to do with school literacy, especially at the secondary school level, the use of standardised measures of literacy has become the norm. According to Amrein and Berliner, if the intended goal of central assessment in the USA is to increase student learning, then, based on the evidence from 18 states, it is not working. In locations where it seems to be working, they argue that the results may reflect coaching or the exclusion of students from the test cohort.

The arguments used to promote central testing are that they:

- (1) tell us what students need to learn and what teachers need to teach;
- (2) make teachers accountable;
- (3) make students work harder; and
- (4) are good measures of the curriculum and student performance.

According to Amrein and Berliner (2002), all of the above statements are ‘likely to be false a good deal of the time’ (p. 5).

Although a range of central assessments were used in the majority of the 60 SSLI schools examined, results from these were not used to compare schools nationally or in any punitive sense, as in the USA. However, international comparisons of literacy achievement made on the basis of central measures including the *Progress in International Reading Literacy Study* (PIRLS) (International Association for the Evaluation of Educational Achievement, 2001) and the *Programme for International Student Assessment* (PISA) (New Zealand Ministry of Education, 2001) have placed literacy and literacy testing into the high-stakes camp. Additionally, the advent of the *National Certificate in Educational Achievement* (NCEA) (New Zealand Ministry of Education, 2002) has brought to New Zealand some sense of the situation pertaining in the USA, although not to the extent that staffing, funding, state control and curricular independence is compromised as pertains there. Results from these senior secondary school examinations, including the number of students achieving NCEA Level 1 English, are published nationally and used to comment on the 'success' of schools. Indeed, some families make important decisions, such as where to live (catchments for schools in New Zealand are zoned), based on the NCEA results as well as on the central assessments that are on offer.

### Five Issues Around Choosing and Using Central Measures

Given the abundance of central assessments used in the secondary schools in the SSLI project and the associated accountability of schools to national benchmarking, the researchers were interested in exploring issues around the selection and use of literacy achievement tests. Based on an analysis of interview data, observations and assessment results from 30 schools sampled on the basis of schools' size, ethnicity, socioeconomic status, type (independent/integrated/state/co-educational/single sex) and regional location, five key issues emerged. These issues centre on: (1) test selection; (2) the time associated with assessment practices; (3) the use of test data; (4) the inappropriate use of tests; and (5) the over-reliance on standardised assessment measures. An over-arching theme relevant to the discussion of all these issues is the progressive modification of assessment protocols and the use of assessment data associated with the SSLI intervention.

#### Test selection

One of the issues that emerged centred on test selection. This is discussed from the perspectives of both teachers and researchers.

##### *Teachers' view*

Based on their observations, media reports and New Zealand Ministry of Education documents, all teachers participating in the SSLI were aware of the wide gap between New Zealand's best and worst readers, and the long 'tail' of poor readers – a feature that is characteristic of New Zealand's results in international literacy surveys such as PISA and PIRLS. They were also aware of the contentious issues around the use of central literacy assessments used nationally and internationally to profile reading achievement. A key issue for these teachers centred on test validity; in the words of one teacher '... they're

not accurate'. The teachers commented that the results did not reflect the reality of students' ability to cope with the literacy demands of their subjects and that the contributing schools' assessment data were 'wrong'. The teachers also mentioned cultural bias, '... the tests don't reflect the world of the students; temporal bias, ... the tests are too old fashioned', and linguistic bias, given the students' preferential use of alternative text forms such as text messages and local discourses. More bluntly, some teachers thought that central assessments were invalid. As one teacher commented in relation to the *Progressive Achievement Tests* (Reid & Elley, 1991), a central measure of reading, vocabulary and listening:

... we didn't trust the data. We've used PATs which is dubious at best. I feel PATs are suspect.

This, together with the comments from other teachers, led the researchers to question: (1) teachers' understanding of the purpose of literacy assessments; (2) their wisdom in comparing assessment results from the instrumentation designed to measure different facets of the literacy construct; (3) their ability to analyse assessment results, and (4) their understanding of the reading process. The implicit view of some of the teachers was that the core constructs underpinning central forms of assessment did not reflect subject-specific literacy demands, or their observations of students' literacy needs. The comments also signal that some schools were at an initial stage in their progression from the use of inappropriate central measures towards the use of both appropriate central and local ecologically valid measures.

#### *Researchers' view*

The research evaluators also wrestled with the issues of test validity when choosing a central measure of literacy achievement that might be administered to provide baseline data in February each year of the SSLI evaluation and re-administered to provide an impact measure at the end of the intervention in November. Given concerns about the validity of generic and uncontextualised literacy assessments, the researchers initially contemplated using the *LanTAC* (Elley, 2003) as a baseline-impact measure. The *LanTAC* is a subject-specific, cloze-format comprehension measure aligned to the key vocabulary of core subjects at Years 9 and 10 (13–14 years old students). It was developed to assist teachers to improve the language and literacy skills of students in all subjects across the curriculum. Each level of the *LanTAC* is made up of two sections: a series of vocabulary tests designed to assess students' understanding of basic terms and concepts used in English, mathematics, science and social studies, and a series of short, cloze comprehension tests based on prose sections from textbooks and other curriculum resources in mathematics, science and social studies. There are three parallel forms of these two sections, each of similar difficulty. Changes in students' *LanTAC* scores from one testing session to another might provide an indication of the magnitude of impact that an intervention program might have had on student achievement.

However, *LanTAC* was subsequently set aside, principally because of a preference among the regional literacy facilitators involved in SSLI for local and contextualised measures that were more sensitive to their specific interventions.

A subsequent compromise reached between the evaluators and the facilitators saw the adoption of the *Essential Skills Assessment: Finding Information in Prose Text* (ESA) (Croft *et al.*, 2000), a 'study skill' type test, as a baseline and impact measure administered to a 25% sample of Year 9 and 10 students in selected case study schools. Although the main purpose of the ESA is formative assessment, it provides three-monthly normed stanines for Year 9 and 10 students. It comprises two sections: Section One measures the students' ability to skim read, and Section Two measures their ability to take notes and comprehend text structure. This central form of assessment is designed to measure skills that are used frequently when secondary school students read for information.

### Tests and time

A second issue to emerge from the SSLI research evaluation into how schools choose and use assessments centred on the *often* inordinate amounts of time that schools spent on administering tests, analysing data, explaining results and distributing test results. In some cases this '... was engaging staff for several weeks, and delaying the distribution of results'. In one school the testing programme extended into the fifth week of the new school year, and the committee formed to process the data did not complete their task until the twelfth week. Some schools used more 'consistent' and 'efficient' ways of administering central measures:

... a hundred boys on their own and all the instructions are taped, so it is very clear. But they are all sitting them [the test] under the same conditions and I feel more confident with that way being done. [Large inner city boys' school]

This siege mentality approach does nothing to establish an environment conducive to 'fair' assessment. Senior management in these schools were aware of the time lag between the administration and the use of analysed assessment data and were keen to further develop their psychometric skills and assessment protocols. The rationalisation of these protocols signalled the progression of schools towards a more manageable regime, and was evident in some SSLI schools at later stages of development (cf. May's discussion of Phase Two and Phase Three schools, this issue).

### Use of test results

The third issue to emerge centred on how schools chose assessment measures and how results were used. Four perspectives are offered: the principals, the teachers, the regional literacy facilitators involved in SSLI and the research evaluators.

#### *Principals*

Principals used test results for a range of purposes that contrasted with those of other users. One principal saw a link between the use of test data (and other forms of qualitative data) and the development of a sharing culture within the school. Specifically, he noted how local data obtained through teachers' 'action research' positively influenced the quality of professional discussions in the



staffroom. In this school, which was moving from Phase Two to Phase Three, the use of local assessment measures was linked to the provision of quality learning by a cohesive community of practitioners.

Another principal with an administrative perspective used graphs of student reading ages, obtained from the assessments administered at the beginning and end of the SSLI intervention, to demonstrate the impact of the SSLI. Other principals used literacy assessment results to monitor policy goals. For example, one principal used the results to indicate whether his school had reached their goal of having 45% of the students reading at or above their chronological age, but then admitted that the goal was 'very, very weak in terms of supporting data'. Another principal stated that he used literacy tests to '... identify intelligent students', although how intelligence was defined was not explained. These principals signal the use of central measures for institutional purposes.

In contrast, other principals admitted that their assessment protocols were light. One stated that the school was '... in the process of having to develop and think about the way we actually work with data', another that he did 'not use much in the way of assessment', and another that tests were used in an 'ad hoc manner' with different departments using different tests. What became clear through these interview data were the idiosyncratic, sometime curious and often policy-orientated ways in which principals used results from central measures of literacy achievement. What also became clear in schools at later stages of development was the progression from using central forms of assessment for institutional purposes, to the more balanced use of central and local measures designed to enhance learning.

### Teachers

Diversity of use and progression were also a defining feature of the ways teachers used assessment data. The researchers established that the majority of secondary school teachers surveyed during the 2003–2005 SSLI period did not use assessment results to inform their practice. Some teachers ignored literacy results because they were yet to see the relevance of literacy to their subject. As one teacher stated, 'Test results are only effective for English teachers because they read them'. Another admitted that test results remain 'in the cupboard'. In one SSLI school, only a third of the teachers accessed results that provided students' reading ages. In another school, it was reported that '... very little was done with the *Probe* (Pool *et al.*, 1999) [test] results other than in the English Department'. During interviews, Heads of Departments commented that they had trouble interpreting test data and this may, in part, account for why evidence about the literacy needs of their students was not informing practice (May, this issue). Some SSLI teachers never progressed from this entrenched position.

Also revealed during interviews was the impact of teachers' resistant disposition. Some felt they had no need to access reading ages because, consistent with the default position in many schools, literacy problems were the responsibility of the English Department, Learning Support Department, Resource Teachers of Learning and Behaviour, Resource Teachers of Literacy, or teacher aides. Indeed, in response to questions about how they would better meet the literacy needs of their students in the future, most senior management staff mentioned that they would be increasing reading specialist hours, thus further removing

the responsibility for literacy from the classroom teacher. It was then unsurprising that in many departments and faculties in these schools, there was a mismatch between students' reading ability and the readability level of the set texts. Thus, the potential benefit of literacy assessment data, indeed the reasons for assessment such as text selection and differentiated instruction, remained unrealised in many cases. While it was true that some teachers used reading ages to select texts and differentiate their instruction, these teachers were, most often, primary trained, with concomitant experience in teaching the reading (and writing) process.

### *Regional literacy facilitators*

Regional literacy facilitators associated with the SSLI also used data (from central assessments, questionnaires and local-teacher-made tests). Initially these data were used to profile literacy achievement at the whole school level, and later used at a class level to initiate discussion about students' needs, and to assist teachers to reflect on the relationships between what the data represented and the design of their programmes. Longitudinally, data were used by facilitators to establish a baseline. Later, it was used to indicate 'where we took them'. The external literacy facilitators' use of literacy test data early in the intervention was revelatory to some teachers. As one teacher noted, albeit in clear deficit terms, '... it [the data] awakened me to the fact that many students are incredibly illiterate'. More positively and developmentally, a regional literacy facilitator recounted the following about a mathematics teacher:

... he came whizzing down when he first got the *Star* (Elley, 2000) results and said, 'I want to show you what I've got. It's really raw data but I'm really excited. The fact that I can see as a statistician, [that] there has been a shift, means I will be carrying out this stuff [using literacy strategies] ... but because we are making a difference, even the likes of cynical me will keep doing it.

That said, some literacy facilitators put assessment results on hold in favour of introducing literacy strategies that teachers wanted, rather than strategies designed to address the evidence-based literacy needs of students. Often, the researchers would encounter examples where the data indicated that students would benefit from comprehension strategies rather than vocabulary strategies. But to engage teachers in the SSLI and to motivate them through short term success, many facilitators would instead resort to modelling vocabulary strategies as their initial preferred professional development option. This was partially because it was at a level at which facilitators could get some level of teacher buy-in. The use of data by regional literacy facilitators also signalled a progression from the use of central data to local data, from a broad institutional focus to a narrower learner focus.

### *Research evaluation team*

The research team used data from central measures of literacy to gauge the sustainability of the SSLI. During 2005, they analysed 2004–2005 data obtained from the administration of the ESA (Croft *et al.*, 2000) to provide a psychometric perspective on the issue of sustainability (Wright *et al.*, 2005b). In one instance

this involved a longitudinal assessment of literacy achievement data from a medium-sized, coeducational, rural school of predominantly indigenous Maori students, located in a low socioeconomic area. Test data were collected from Year 10 students (who were first tested as Year 9 students in 2004), and Year 11 students (who were first tested as Year 10 students in 2004). Raw scores from 17 of the 25 Year 10 students who were re-administered the ESA in October 2004 were subjected to a paired sample *t*-test to investigate any association between the SSLI intervention period and their test scores. There was a statistically significant increase in ESA raw scores from October 2004 ( $M = 19.12$ ,  $SD = 3.87$ ) to April 2005 ( $M = 21.88$ ,  $SD = 3.04$ ,  $t(16) = 2.63$ ,  $p = 0.02$ ). The eta-squared statistic (0.30) indicated a large effect size associated with the SSLI intervention. These results also indicated a significant increase in mean total raw scores and a narrowing of the spread of scores during the period of the SSLI.

Fifteen of the seventeen Year 11 students who were re-administered the ESA as Year 10 students in October 2004 remained at school and were re-tested at the end of April 2005. There was a statistically significant increase in ESA scores from October 2004 ( $M = 17.80$ ,  $SD = 3.16$ ) to April 2005 ( $M = 20.46$ ,  $SD = 1.92$ ,  $t(14) = 3.0$ ,  $p < 0.01$ ). Again, the eta-squared statistic (0.39) indicated a large effect size.

These data were used by the researchers to suggest that pedagogical and policy changes associated with the SSLI were sustaining literacy achievement at this school 18 months after the commencement of the intervention. Principals, teachers, regional literacy facilitators and researchers used literacy assessment data in diverse ways. What emerged strongly from the research was that teachers' initial literacy practices in many instances were not particularly evidence-based, but that over time there was a gradual progression, associated with an acceptance that literacy was every teacher's responsibility, towards the greater use of literacy assessment data as a basis for teaching and learning.

### Inappropriate tests used inappropriately

The fourth issue to emerge with respect to choosing and using literacy assessments centred on whether the measures administered were 'appropriate' and whether the data from these measures were used appropriately. For example, one school administered a bank of tests as 'tools for diagnostic purposes', a purpose for which the tests were not designed. Another school, with a high proportion of Maori and Pasifika (Pacific Island) students used the *Gapadol* (McLeod & Anderson, 1973) test, a cloze-type reading test normed in the USA, that was both culturally and psychometrically inappropriate. Another school used the *Star* (Elley, 2000) literacy test with an age group for which it was not designed, justified on the basis that the students were not capable of completing the age appropriate test. Another school published *Burt* (Gilmore *et al.*, 1981) word recognition scores as reading comprehension ages. While the variety of assessments used were not necessarily problematic, it is of concern when measures are used inappropriately and when their use is coupled with an incomplete understanding about the purposes of the measures.

Similar issues of inappropriateness emerged around the use of local teacher-made assessments. Concerns about the construct validity of teacher-made

literacy assessments were acutely exemplified by a literacy 'entry examination' constructed by an English Department at a large single-sex school in an upper socioeconomic area of a large urban centre. This entry examination was administered to all new Year 9 students (approximately 13-year-olds) as a means for streaming or tracking them into academic class groups. According to Limbrick *et al.* (2003), any definition of literacy should focus on the construction of meaning. Construction implies language in use and use implies thinking. Others define literacy as a social practice (Hull, 2003; May, this issue) involving oral, written and visual tools. Clearly, literacy is more than spelling and grammar, but the 'entry examination' comprised of: (1) 45 vocabulary and comprehension items; (2) language skills (40% of the test) that included items on subject, predicate, verbs, nouns, homophones and proof reading, and (3) a creative writing item that instructed the students to '... write in paragraphs and use adjectives, verbs and nouns' and '... be imaginative', and whose assessment criteria was '... grammar, spelling, punctuation, structure and the development of ideas'. Measured against contemporary definitions of literacy, the entry examination reflected a narrow interpretation of the construct, and called into question the validity and appropriateness of the examination and its subsequent use as a mechanism for streaming or tracking Year 9 students into 'ability grouped' classes. Many SSLI schools maintained their use of literacy data for streaming, while realising the potential of other diagnostic and local measures that might inform pedagogy.

### Over-reliance on standardised measures

The final issue to emerge in respect to choosing and using assessments concerned the over-reliance on, and misplaced belief in the rigour of central measures of literacy achievement. For example, there was a widespread belief among senior management in most schools involved in the SSLI that the best way to measure the impact of the intervention was to administer a central measure. Typical of this belief was a comment that:

I cannot comment [about changes in students' literacy achievement during the SSLI] yet because the *Star* [a standardised test of reading achievement] testing [results] is not processed yet.

Belief in the rigour of central assessments was captured in the comments of one teacher who stated that 'the impact of the SSLI would be seen in *National Certificate of Educational Achievement* (NCEA) results' – New Zealand's standards-based secondary school examination. This teacher saw NCEA as an indicator of progress, and saw the use of literacy strategies introduced through the SSLI as impacting on NCEA scores:

I have had students ask me to repeat literacy strategies ... because they found it improved their ability to answer Level One and Level Three NCEA Physical Education tasks ... it [literacy strategies] improved students' ability to analyse NCEA questions.

Other teachers also acknowledged that the potentially high-stakes profile of the NCEA was a driver in respect of their involvement in the SSLI. They

reported helping students prepare for examinations (that is teaching-to-the-test) by concentrating on literacy strategies that assisted students to unpack examination questions. This behaviour is reminiscent of the type of coaching that Amrein and Berliner (2002) criticise in relation to high-stakes testing in the USA. This deference to the power of central assessments was also seen in the way some regional literacy facilitators measured the impact on literacy achievement associated with their interventions. Some facilitators reported the data obtained from parallel test forms administered at the beginning and end of the intervention. Others who did not, drew the ire of principals.

Not having the analysed feedback [from the facilitator] . . . of our students' development between the initial test and final test was one of the least helpful aspects of the SSLI project. [Principal]

It would appear that in many secondary schools, over-reliance on central assessment measures is, overwhelmingly, driven by the knowledge-as-object metaphor; not by concerns about learning or knowing, not by the use of assessments sensitive to the literacy demands of each subject or the literacy and learning needs of students. Together, the evidence around these five 'choose and use' issues suggest that many schools were guilty of excessive testing, of under-analysing data, and of choosing and using assessments without any clear rationale other than historic precedent. It also signalled the progress of some schools around these issues as a result of their involvement in the SSLI intervention. This progress was seen most clearly in the use of local assessment measures. These measures were seen as providing an additional means of assessment *for* learning and assessment *as* learning.

### Case Study: Ecologically Valid Assessment Practices

As already argued, there can be tensions between teachers' use of local, subject-specific assessments, and institutional preferences for central assessments. Likewise, there can be tensions around the appropriateness of different forms of local assessment. For example, while local assessments might reflect what was taught, they might lack *ecological validity*, that is, failure to reflect the literacy and thinking tools used.

In psychometric terms, the concept of ecological validity derives from Neisser's (1976) argument that cognitive research is not representative of real-world situations in which people think and act. Central to Neisser's argument were issues around the extent to which laboratory findings generalise and transfer to real life settings. Invariably, psychological research and literacy assessment practices involve a trade-off between central control, exemplified by standardised measures, and forms of local control exemplified by ecologically valid measures. In the classroom context, the key characteristic of an ecologically valid form of assessment is that it should reflect the 'conditions' that pertained when the content to be assessed was learnt. In the literacy context, this means that assessments should reflect the literacy and thinking tools used by teachers and students when the content was taught and learnt. In a social context, the use of ecologically valid forms of assessment that reflect the use of literacy and thinking tools is an inclusive literacy event. These can be contrasted with the use of

central measures that can often be characterised as exclusive and discriminatory (Corson, 1999).

The following case study describes the experiences of seven participants teaching Year 9 and 10 English, health education, biology and physics in four of the SSLI 2003–2005 case study schools.

### Participants

The motivation of the seven participants involved in this case study stemmed from their concerns: (1) around the over-testing and under-analysing of data obtained from a range of central literacy measures (Whitehead *et al.*, 2004) and (2) that historic unit assessments failed to reflect their recent use of the literacy and thinking tools. These seven teachers had been identified by the school literacy leaders as being actively engaged in the SSLI. An analysis of their historic unit assessments revealed that they possessed a high degree of content validity, i.e., they rigorously reflected the specific domain of the content taught (Carmines & Zeller, 1991). The teachers were allocated time with the author to: (1) discuss the literacy challenges of the texts they were using; (2) review a set of text-linked, subject-specific literacy and thinking tools (Whitehead, 2006); (3) discuss how these tools might be reflected in the design of unit tests, and (4) select tools appropriate to the texts and tasks that they felt confident in using. Through discussion, the teachers reasoned the following.

- (1) A pre- and post-tests design would allow them to measure changes in student learning, and forestall the need for parallel unit test forms.
- (2) Pre-tests would provide them with a measure of judging students' prior knowledge.
- (3) An ecologically valid assessment would reflect their use of literacy and thinking tools.
- (4) Results could be compared to those from historical measures for the same unit.

Teachers then redesigned their historic unit assessments to reflect the academic literacy demands associated with learning that content, and their use of literacy and thinking tools. In respect of academic literacy demands, the teachers had observed that students were not coping with subject-specific vocabulary. As Corson (1999) notes '... the importance of words in education seems so obvious that it was taken for granted for much of the history of schools throughout education' (p. 21). But the vocabulary of academic texts can be challenging, especially if writers assume that student readers have sufficient prior knowledge to successfully decode and comprehend technical vocabulary without the support of a rich context. Given Olson's (1997) contention that oral participation is a key to vocabulary growth, and given that academic vocabulary provides the conceptual links for learning, it was unsurprising that the seven teachers opted to use oral vocabulary learning tools.

### Test items

Given their students' diverse academic vocabulary needs, seven Year 9 and 10 teachers used a *Vocabulary Matching* tool described in *Effective Literacy*

*Strategies in Years 9–13* (Ministry of Education, 2004) and a *Concept Frame* tool (Whitehead, 2001). Only one teacher, who chose to revert to pre-intervention transmission pedagogy after a few weeks because ‘the class was out of control’, withdrew from the case study. The students first encountered the Vocabulary Matching and Concept Frame tools in a unit pre-test. They encountered the two tools again during lessons, and finally in the form of post-test items (see Figures 1 and 2). Consistent with the concept of ecological validity, some of the assessment items in each subject area reflected the use of these tools.

By way of example, the Vocabulary Matching and Concept Frame tools used by the English teachers (see Figures 1 and 2) assisted students to acquire vocabulary that described the characteristics of narrative participants.

*English Test Vocabulary Matching Item*

Instructions: Match the words in the left column with their meanings in the right column.  
Write your answers in the box below.

Characteristic		Reason	
1. Caring		A. He took only one necklace	
2. Kind		B. She cooked more rice balls for him after the first ones rolled away.	
3. Mean		C. He took all the jewels	
4. Greedy		D. He tricked the mouse	
1	2	3	4

**Figure 1** English Test Vocabulary Matching Item

*Concept Frame Item*

Instructions: Use the *Concept Frame* to help you describe your character.

My character is (is a...)....	My character can...
1	1
2	2
3	3
Examples of other characters like my character are...	My character has (has a...) ....
1	1
2	2
3	3

**Figure 2** Concept Frame Item

### Students' response

The students' response to the pre-test and post-test and to the teachers' use of the tools was positive. The pre-test provided students with clear goals, as one student's stated: '... [it was] very clear what it was we had to learn and work on improving'. Students' response to the post-test item was also positive because it indicated to them whether they had achieved their goals. As one student noted '... it showed us what we had learnt'.

Beyond novelty value, the Vocabulary Matching and Concept Frame tools were valued by students because their use was associated with improved assessment scores, especially among less able students. As one teacher commented '... many of my low-middle ability students did make good improvements (5–10 marks)'. Another noted that:

The one student who has the greatest learning difficulties did score fairly highly in the unit post-test, achieving 40% on one section, which for her is a great achievement.

In addition to improved scores, teachers who used the Vocabulary Matching and Concept Frame tools reported high levels of engagement among students and rapid acquisition of subject specific vocabulary.

### Teachers' response

The use of the pre-post test format, the two teaching tools, and the redesigned unit assessments, translated into positive attitudes among the teachers. For example, Year 9 and 10 English teachers stated that the administration of the pre-test had made it easier for them to identify the vocabulary needs of students. Indeed, they had decided to use the same pre-test and post-test procedure for their next junior writing unit and extend the use of the tools and the design of ecologically valid assessments into the senior school. Although the use of the two tools and the redesign of historic unit assessments was a significant risk for the teachers, it was deemed as a risk well taken. The risk was taken because teachers became confident that they could use the tools, because the culture within the school, and specifically the school's literacy leader (Wright, this issue), was supportive, and because most of the teachers had progressed to a phase of development characterised by the consolidation of effective literacy practices (May, this issue; Wright *et al.*, 2005a).

However, for a newly graduated physics teacher in a school with a transmission teaching culture, the use of vocabulary teaching tools and their inclusion as ecologically valid assessment items posed a challenge:

As a first year teacher still coming to grips with the culture of learning at the school – and with the way science is currently taught in the school, I did find it difficult to introduce some things [literacy strategies] .... The students have developed a culture of demanding to be spoon-fed information and so attempting group exercises is often difficult.

On the contrary, some teachers rated the process of using the Vocabulary Matching and Concept Frame tools, designing ecologically valid test items, and administering pre- and post-tests as a success. As one teacher put it, 'I



have found the exercise very useful, particularly in terms of reliable assessment practice linked to deliberate acts of teaching.' The redesigned unit assessments allowed teachers to measure students' prior knowledge, helped their students develop anticipation sets for the academic vocabulary associated with the content of the units, and measured learning.

This case study illustrates the use of local assessment practices that, the teachers felt, surpassed the value of their historic, content-focussed unit assessments. The case-study teachers courageously operated without an authoritative set of mandated core literacy and thinking tools, against which the performance of schools might be compared nationally. The local assessments designed by these teachers' privileged professional choice and reflected student needs and teacher confidence over externally imposed central assessment protocols.

## The Way Forward

The central and local forms of assessment that teachers use have a powerful influence on the kinds of teaching the students encounter, and the kind of learning students accomplish. There is nothing inherently wrong with the use of central measures of literacy and learning for comparative purposes. Equally, there is nothing inherently wrong with local subject specific assessments. Both have their specific roles. But the validity of both types of assessment can be improved if they have a degree of *ecological validity*, that is, if they reflect how teachers teach, how students learn, and how students use the content assessed. From a teacher's perspective, the use of ecologically valid assessments is about the positive effects of assessing both what was taught (knowledge-as-object) and students' use of literacy and thinking tools as ways of knowing. From a student's perspective, this form of assessment acknowledges the congruency between what was learnt, the tools used to learn it and the design of the assessment measure. Teachers in the case study just described were able to enhance the ecological validity of their assessments.

## From knowledge testing to knowing

The way assessments were used in the case study seemed to mark developmental phases on two continua associated with the transition from the central to local dominance of assessment measures. One 'assessment' continuum (Wright *et al.*, 2005b) seems to begin with understandings around the need to use literacy data and to change data collection procedures. These emerging understandings lead to opportunities, in association with the mentoring role of external literacy facilitators and school-based literacy leaders, for schools to deepen their understandings about how to use appropriate tests in appropriate ways. Change is then realised through the implementation of new forms of assessment, and will probably conclude with a period of reflection and adaptation, leading to embedded praxis (Jackson, 2004; May, this issue).

Along a second 'teaching and learning' continuum, the use of ecologically valid assessment measures seem to mark a Kuhnian-type shift in thinking from a knowledge-as-object, mind-as-container metaphor, towards the adoption of a knowledge-as-event metaphor. The former metaphor fails when assessment is constructed in terms of helping people learn how to do things, and when

learning is seen as a knowledge-building enterprise facilitated by a set of core literacy and thinking tools that develop students' capacity for knowing. Central and local literacy and subject-specific assessments tend to facilitate the knowledge-as-object, mind-as-container metaphor and fail to capture any sense of how knowledge works in English, science or any other subject. These assessments, used widely in SSLI schools, are content-focussed rather than system-focussed and force students to become consumers of content rather than practitioners of process. These assessments are also aligned with a transmission approach to teaching.

The progress made by the teachers reported in the case study suggests that current assessment practices in literacy and secondary school subjects are not ambitious enough. They fail to reflect that knowledge is no longer only a quantifiable (and therefore easily assessed) 'thing' organised into subject disciplines, but rather an event that does things or causes things to happen (Castells, 2000). If we accept that knowledge has what Lyotard (1984) calls *performativity*, it follows that assessments of literacy and curriculum subjects should measure, at a local level, what the knowledge of these do and the tools used to manipulate that knowledge. This type of assessment is consistent with the value that society now places on the ability to produce new knowledge rather than consume old knowledge. As likely elsewhere, current central assessment protocols in New Zealand secondary schools are often inconsistent with future thinking, encapsulated by Claxton's (2002) 'school as gymnasias', 'fit mind', and 'mental exercise' metaphors. But when we work out at the gymnasium, we don't look at the equipment, we use it. Likewise, assessment should tell us how well a student can use what they know.

## Conclusion

The central and local assessment measures used in many of the SSLI schools rarely assessed students' ability to create knowledge. Rather, they reflected a conception of knowledge as a noun, a thing or a commodity. There is evidence that some schools selected inappropriate measures and that some schools were over-reliant on these to 'sort out' students. There appears to be no social justice in the use of assessments that make all students measure up to the preset norms of specific discourses and common identities that reflect the cultural capital of an industrial age ruling class. The choice of these traditional assessments and site specific content-focussed unit tests embody the cultural capital (Bourdieu, 1973) of the old ruling class, and suggest teachers' choice of assessments is consistent with the knowledge-as-object, mind as mind-as-container metaphor.

In contrast, the progression of some schools towards the complementary use of central and ecologically valid assessments is consistent with both identifiable phases of school-based literacy development and with the new cultural capital of an age in which knowledge is defined through what it can do (May, this issue). The appropriate balance between local and central assessment measures, between content and systems perspectives, may be difficult to maintain. However, it reflects the importance of a teacher's decision to move from being merely a transmitter of information, who ignores the literacy profiles of their students, to being a co-constructor of a literate society.

## Correspondence

Any correspondence should be directed to Dr D. Whitehead, University of Waikato, School of Education, Private Bag 3105, New Zealand (davidw@waikato.ac.nz).

## Notes

1. Opinions expressed in this article are those of the author and do not necessarily coincide with those of the New Zealand Ministry of Education.
2. The New Zealand school year runs from January to December.

## References

- Amrein, A. and Berliner, D. (2002) High-stakes testing, uncertainty and student learning. *Education, Policy Analysis Archives* 10 (8). <http://epaa.edu/epaa/v10n18/>. Accessed 15.11.06.
- 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* (pp. 137–149). Toronto: University of Toronto Press.
- Black, P. and Wiliam, D. (1998) Assessment and classroom learning. *Assessment in Education* 5 (1), 7–74.
- Bourdieu, P. (1973) Cultural reproduction and social reproduction. In R. Brown (ed.) *Knowledge, Education and Social Change* (pp. 71–112). London: Tavistock.
- Brock-Utne, B. (1996) Reliability and validity in qualitative research within education in Africa. *International Review of Education* 42 (6), 605–621.
- Carmines, E. G. and Zeller, R.A. (1991) *Reliability and Validity Assessment*. Newbury Park, CA: Sage.
- Castells, M. (2000) *The Rise of the Network Society* (2nd edn). Oxford: Blackwell.
- Chaytor, N. and Schmitter-Edgecombe M. (2003) The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review* 13 (4), 181–97.
- Cicourel, A.V. (1996) Ecological validity and ‘white room effects’: The interaction of cognitive and cultural models in the pragmatic analysis of elicited narratives from children. *Pragmatics and Cognition* 4 (2), 221–264.
- Clarke, S. (2001) *Unlocking Formative Assessment*. London: Hodder & Stoughton.
- Claxton, G. (2002) *Building Learning Power: How to Help Young People Become Better Learners*. Bristol: TLO.
- Clay, M. (2000) *Running Records for Classroom Teachers*. New York: Heinemann.
- Corson, D. (1999) *Language Policy in School*. Mahwah, NJ: Lawrence Erlbaum.
- Croft, C., Dunn, K. and Brown, G. (2000) *Essential Skills Assessment: Finding Information in Prose Text*. Wellington, NZ: New Zealand Council for Educational Research.
- Elley, W. (2000) *STAR Supplementary Test of Achievement in Reading Years 3, 4-6 & 7-9*. Wellington, NZ: New Zealand Council of Educational Research.
- Elley, W. (2003) *LanTAC: Language Tests Across the Curriculum*. Thames: Thames High School.
- Freebody, P. and Wyatt-Smith, C. (2004) The assessment of literacy: Working the zone between ‘system’ and ‘site’ validity. *Journal of Educational Enquiry* 5 (2), 30–49.
- Garfinkel, H. (1967) *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.
- Gilbert, J. (2005) *Catching the Knowledge Wave?* Wellington, NZ: New Zealand Council for Educational Research.
- Gilmore, A., Croft, A. and Reid, N. (1981) *Burt Word Reading Test – New Zealand Revision*. Wellington, NZ: New Zealand Council for Educational Research.
- Higginson, C.I., Arnett, P.A. and Voss, W.D. (2000) The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Archives of Clinical Neuropsychology* 15 (3), 185–204.

- Hull, G. (2003) Multiple media and agentive selves: Adults and youth craft digital stories. A keynote address presented to the Language, Education and Diversity Conference, University of Waikato, Hamilton, 26–29 November.
- International Association for the Evaluation of Educational Achievement (2001) *Progress in International Reading Literacy Study*. Amsterdam: IAE.
- Jackson, N. (2004) Developing the concept of metalearning. *Innovations in Education and Teaching International* 32 (4), 391–403.
- Limbrick, L., Parkhill, F. and Smith, J. (2003) Developing a qualification for specialist literacy teachers. *English in Aotearoa* 51, 22–29.
- Lyotard, J.F. (1984) *The Postmodern Condition: A Report on Knowledge*. Manchester: Manchester University Press.
- McLeod, J. and Anderson, J. (1973) *Gapadol Reading Comprehension*. Oxford: Heinemann.
- Ministry of Education (2001) *Programme for International Student Assessment*. <http://www.minedu.govt.nz/goto/pisa>. Accessed 15.11.06.
- Ministry of Education (2002) *National Certificate in Educational Achievement*. Wellington, NZ: Learning Media.
- Ministry of Education (2004) *Effective Literacy Strategies in Years 9–13*. Wellington, NZ: Learning Media.
- Mook, D.G. (1983) In defence of external invalidity. *American Psychologist* 38, 379–87.
- Neisser, U. (1976) *Cognitive Psychology*. New York: Appleton-Century-Crofts.
- New Zealand Council for Educational Research (2001) *Progressive Achievement Test of Reading Comprehension*. Wellington, NZ: New Zealand Council for Educational Research.
- New Zealand Council for Educational Research (2001) *Essential Skills Assessment: Finding Information in Prose Text*. Wellington, NZ: New Zealand Council for Educational Research.
- Olson, D. (1997) Talking about text and the culture of literacy. In B. Davies and D. Corson (eds) *Oral Discourse and Education* (pp. 2–9). Boston: Kluwer.
- Pool, B., Parkin, C. and Parkin, C. (1999) *Prose Reading Observation, Behaviour and Evaluation of Comprehension*. Wellington, NZ: Triune Publications.
- Reid, N. and Elley, W. (1990) *Progressive Achievement Test: Vocabulary* (Revised). Wellington, NZ: New Zealand Council for Educational Achievement.
- Reid, N.A. and Elley, W.B. (1991) *Progressive Achievement Tests Teacher's Manual: Reading Comprehension, Reading Vocabulary*. Wellington, NZ: New Zealand Council for Educational Research.
- Scholastic (2006) *Scholastic Reading Inventory*. New York: Scholastic.
- Whitehead, D. (2001) *Top Tools for Literacy and Learning*. Auckland, NZ: Pearson Education.
- Whitehead, D. (2004) *Top Tools for Teaching Thinking*. Auckland, NZ: Pearson Education.
- Whitehead, D. (2005) Testing like you teach: Effective literacy strategies in the design of unit assessments. *English in Aotearoa* 56, 3–12.
- Whitehead, D. (2006) Justifying what we do. *English in Aotearoa* 57, 3–12.
- Whitehead, D., May, S. and Wright, N. (2004) *Secondary Schools' Literacy Initiative (SSLI) Research Evaluation: Literacy Assessment and Achievement, 2003–2004*. Hamilton, NZ: Wilf Malcolm Institute of Educational Research, University of Waikato, for the Ministry of Education.
- Wright, N., Smyth, J., May, S., Whitehead, D. and Smyth, S. (2005a) *Secondary Schools Literacy Initiative Research Evaluation. Report on 2004 Case Study Schools*. Hamilton, NZ: Wilf Malcolm Institute of Educational Research, University of Waikato.
- Wright, N., May, S., Whitehead, D., Smyth, J. and Smyth, S. (2005b) *Secondary Schools Literacy Initiative Research Evaluation. Sustaining Literacy Initiatives in Secondary Schools*. Hamilton, NZ: Wilf Malcolm Institute of Educational Research, University of Waikato.