English Teaching: Practice and Critique December, 2008, Volume 7, Number 3
http://education.waikato.ac.nz/research/files/etpc/2008v7n3art2.pdf pp. 10-25

Testing like you teach: The challenge of constructing local, ecologically valid tests

DAVID WHITEHEAD
University of Waikato

ABSTRACT: In an educational context, local, ecologically valid tests can reflect the use of literacy and thinking tools. These tests present a challenge to central, content-focused, high-stakes testing, and to transmission approaches to teaching. They require teachers to accept knowledge as a verb, and to design assessment protocols that reflect co-constructive ways of teaching. This article reports the outcome of praxis action research with middle and secondary school teachers who incorporated topic-appropriate literacy and thinking tools into their teaching. They also redesigned their local tests linked to high-stakes test protocols to reflect the use of these tools. A thematic analysis of observations and interviews suggests that this process impacted on the structural characteristics (morés) of the schools, and posed affective, cognitive and pedagogical challenges to teachers.

KEYWORDS: Ecologically valid, tests, literacy, teaching

PROLOGUE

There was one thing that, as a secondary school student, I could guarantee would happen in February. While summer holiday memories of bleached whalebones and bronzed bodies in bikinis were still fresh in my adolescent mind, I could guarantee there would be central, standardized tests of vocabulary and reading comprehension. My sensory memory is still filled with the smell of waxed floors, and sweaty bodies too big for last year"s uniforms, seated in the hall, in rows – this was mass testing in a Catholic school, and it was awful. I was never told that most secret statistic, my class percentile score, although it was once shared with my parents who no doubt nodded with sincere misunderstanding.

INTRODUCTION

Summative tests of literacy achievement form a dominant feature of the pre-tertiary educational landscape in most "developed" countries. These tests take three main forms: (i) international, norm-referenced "high-stakes" tests including PISA (Organization for Economic Cooperation and Development, 2006), and PIRLS (International Association for the Evaluation of Educational Achievement, 2006), (ii) national norm-referenced measures in the form of age and grade-targeted measures used to compare literacy achievement, and (iii) site specific or "internals". While governments control decisions around the administration of international measures, the *Secondary School Literacy Initiative* (SSLI) research and evaluation project 2003–2005 (Wright, May, Whitehead, & Smyth, 2005) in New Zealand revealed that there had been a rationalization in the number of national,

summative/diagnostic, norm-referenced measures administered in the period of the project. For example, in 2003 some 53 different tests were used in the 60 secondary schools surveyed, but by 2005 this had fallen to 23. Since 2005 this has fallen again to around 10 as the highly funded Ministry of Education options of asTTLe Reading and asTTLe Writing (Ministry of Education & University of Auckland, 2003) have been "rolled out" into schools. The use of data from these tests by senior management teams remains, primarily, to stream/band students to allocate resources. As yet, the use of national and site-specific test data to inform teaching is minimal.

The researchers also found that the selection of subject-specific, site-specific summative tests, including those used by teachers of English, tended to be based on historic precedent. The construction of these tests reflected the design and content of high-stakes measures administered to 15-18 year old students. Exceptions were found in the few teachers engaged in the SSLI, who reflected their classroom use of effective literacy strategies in the construction of summative "topic" tests. These strategies had been introduced by SSLI regional facilitators and in-school literacy leaders. This article looks at the efforts of these teachers, in the context of a national assessment system that influences what teachers teach, how they teach, and how they test.

The claim that central, high-stakes tests of literacy achievement administered under controlled conditions are unrepresentative of socially contextualised literacy tasks that occur in dynamic instructional settings is not new (Brock-Utne, 1996; Freebody & Wyatt-Smith, 2004; Neisser, 1976). Indeed, it is questionable whether these central forms of literacy assessment either reflect how students become independent literate thinkers, or predict their success on literacy tasks (Higginson *et al*., 2000). In a culture of high-stakes, content-focused, central and local tests, the use of alternative tests designed to measure what students can do with what they know can be easily marginalized (Amrein & Berliner, 2002; Cizek, 2001; Gambell, 2004).

Despite these claims, the administrative cudgel of schools who inappropriately administer standardized, summative (rather than potentially more useful diagnostic) tests of literacy (and virtually no tests of thinking), and who use the results to stream classes, has become an all-too-memorable rite of passage for many students commencing secondary school. In English and other subjects, these tests initiate students into a pervasive culture of assessment linked to national curriculum standards. This culture of high-stakes testing accepts examination prescriptions (The *National Certificate of Education Achievement* in New Zealand) as the default curriculum.

These tests rarely reflect how expert teachers of English address the literacy and thinking demands of their subject, and rarely mimic the dynamic processes of teaching and learning that pertained when students came to know that subject (Whitehead, 2005; 2006). Rather than possessing the *site-validity* of local assessment measures, these examination prescription-linked central measures possess *system-validity,* because they reflect the needs of policy-makers and bureaucrats to profile student achievement. The reality is that in New Zealand and elsewhere, central assessments of English coupled with examination prescriptions have a powerful influence on what teachers teach, on how teachers teach and on how students learn.

While these tests might provide some measure of achievement, they rarely reflect the pedagogical tools teachers use to facilitate learning.

BACKGROUND TO THE ACTION RESEARCH

The praxis action research reported here was initiated within middle and secondary school sites. These sites were characterised by the use of high-stakes, central tests, and local, content-focused, summative assessments of English and other subjects. The purpose of the research was to observe and measure the effect of introducing literacy and thinking tools as pedagogical aids, and potential assessment items, within these sites – a purpose that might be seen as challenging the status quo.

Based on an inspection of local tests used historically at these sites, and discussion with teachers, it was evident that their tests reflected a transmission-of-content approach to teaching, because most test items asked students to recall content conveyed by the teacher, text or web-site. Although these measures had content validity, they did not reflect the use of literacy and thinking tools, either because these were not part of teachers" pedagogical content knowledge, or, if used, were not valued sufficiently to be reflected in test design.

The researcher's challenge to the English teachers" pedagogical content knowledge, and to these historic measures, had its genesis in two events. The first event was a series of adverse school reviews from an external agency that characterised the pedagogy used in these inner city, elite, high socio-economic catchment, boys' school as "wanting", especially in respect to improving students' literacy and higher order thinking. Coupled with these reviews was the pressure to compete in state examinations with similar élite schools. The second event was an invitation from both schools to the researcher to visit and engage in some praxis action research, and an associated expression of interest from volunteer teachers who were willing to take the risk of participating in research designed to change their pedagogy and assessment practices. The praxis action research took place over a three-month period.

SITUATING THE RESEARCH

A praxis action research paradigm

The action research was situated within a praxis research paradigm that shares a number of perspectives in common with the interpretive action research paradigm (Lather, 1986). Praxis (Aristotle) is the art of acting upon the conditions one faces in order to change them. In the context of this research, these were in the form of changes to pedagogical and testing practices. This paradigm rejects the notion of researcher neutrality, understanding instead that an active researcher is often one who has most at stake in resolving a problematic situation. Consequently, this paradigm accommodates the role of researcher as a pedagogical change-agent, focused on empowering teachers to become better practitioners.

The praxis action research paradigm, or collaborative inquiry paradigm, was chosen because it accommodated the involvement of volunteer teachers who, through self-reflection, had identified problems with their pedagogical and assessment practices, and who had decided to

address those problems with the help of the researcher. The paradigm accommodates, on an ongoing and cyclical basis, the plan, action, observation and reflection cycle typical of action research (Kemmis, 1988). A key aspect of the reflection component in the present research was post-lesson debriefings/semi-structured interviews and written feedback that prompted teachers to consider their practice and the implications of that practice for assessment.

A psychometric paradigm

The action research was also broadly situated within an empirical psychometric paradigm. Specifically, it was situated within a paradigm underpinned by the belief that assessment can be designed to reflect the characteristics of classroom pedagogical interactions. This belief is captured in the concepts *ecologically validity* and *cognitive interdependence.* Together these concepts suggest students should encounter the same teaching and learning processes in their assessments as they encountered when learning the content; that teachers should test like they teach.

The concept of *ecological validity* derives from the more commonly used and broader concept of *test validity*. Specifically, *ecological validity* derives from Neisser's (1976) and later Haber's (1983) and Gibson's (1997) contention that experimental psychology is not representative of real-world situations in which people think and act. Consistent with this argument, the key principle underpinning the definition of an *ecologically valid* assessment is that students must react to the test, to some extent, as if they were participating in an instructional situation (Cicourel, 1996). In the context of this research, this contention can be restated as: English assessments are ecologically valid if they reflect the use of literacy and thinking tools used to help students learn and become independent literate thinkers.

Methodology

Before outlining the procedures adopted for this research, it is worth noting factors that worked against the researcher's acceptance within these sites (Schensul, Schensul, & Lecompte, 1999). One factor was the "unknown". The teachers at these schools had no prior knowledge of the researcher, although the Deans in each school had the researcher's CV. Acceptance among teachers of this "unknown" researcher was earned through demonstrations that he "knew his stuff" (based in part on presentations to senior staff that focused on a justification for the use of literacy and thinking tools as assessment devices), and that the researcher could "do his stuff" (based in part on the researcher working with students while teachers, who had volunteered to participate in the research, observed).

A second factor that inhibited the researcher's acceptance related to the schools' structural characteristics, that is, to the morés that existed in the schools around pedagogy and assessment. The schools were characterised by traditional, transmission-type approaches to teaching and the use of local content-focused summative tests. The introduction of literacy and thinking tools as pedagogical aides and potentially, as test items, was seen by some teachers as a challenge to this structure.

A third factor working against the researcher's acceptance was his shared preference that knowledge should be re-defined as a verb. Given that many teachers were tied to

"knowledge-as-object", "mind-as-container" metaphors, they became resistant to this redefinition of knowledge. The view of knowledge as an event that does things or causes things to happen (Castells, 2000) challenged many teachers. The researcher guided the volunteers, and by association other staff, into accepting that knowledge has what Lyotard (1984) calls *performativity* and, consequently, that their teaching and testing should focus on the use of what the students knew. The rationale for this position was that teaching and testing should be consistent with the value society now places on the ability of learners to produce new knowledge, rather than consume and reproduce old knowledge. This position was a direct challenge to the existing structural characteristics of the schools, and especially their assessment protocol, and comments made to both the volunteers, and the researcher from some staff indicated that they were resistant to this new definition of knowledge – without consideration of whether their resistance might be linked to issues of social justice.

Although there were factors that militated against the researcher's acceptance, these were soon dispelled among the volunteer group, as signalled by teachers" inviting the researcher to numerous social and sporting events, and invitations from teachers from a "safe distance" outside the volunteer group, especially from senior teachers, to discuss their teaching and testing protocols. It was also dispelled by the researcher's conscious use of jargon-free language and the message to volunteers from the outset that they governed what did or did not happen (deWalt & deWalt, 2002). The praxis action research paradigm adopted by the researcher was never likely to work unless there was "buy-in" from the volunteers. Although there were various impediments to acceptance, a group of six teachers volunteered to work with the researcher.

Procedures

The procedure for obtaining participants was to first outline the purpose of the research to senior management teams. Key points presented during this briefing were that:

- The research was aimed at obtaining rich case studies that described the work of teachers, who would be invited to use a selection of literacy and thinking tools for teaching, learning and assessment;
- All teachers would need to be volunteers;
- Tools introduced would aligned with teachers" current topics, and would be selected and modelled to each participant;
- Modelling would be, most often, undertaken with the teacher and the researcher, but might be followed by the researcher working with a group of students while the teacher observed;
- The researcher would observe volunteer teachers as they used tools in the context of a lesson, and would de-brief the teacher using a semi-structured interview that included the following guide questions:
    - Did the use of the tool have any effect on the way you taught?
    - Were there any differences from previous lessons in the way students responded to you, and your use of the tool?
    - If you taught the lesson again would you want to change anything?

    - Where to from here?
- Teachers would discuss with the researcher how the literacy and thinking tools might be used as assessment items. (This included the researcher describing the concept of *ecological validity* and commenting on tests associated with the topic used historically);
- Teachers would receive written feedback describing what the researcher saw, and suggestions for the teacher in regard to their future use of tools;
- Feedback from teachers during interviews and observation notes would be examined for themes.

Characteristics of volunteers

There were two characteristics common to all volunteers. The first was a subscription to the concept of "best practice". This characteristic was articulated diversely. For example, one teacher, a recent graduate, who was facing classroom management and curriculum content concerns, was motivated to volunteer because she thought the researcher might offer something to address these concerns. A second teacher "controlled" his class by adopting a transmission approach to teaching and an authoritarian approach to classroom management. While he was "coping", he looked to the researcher for ways of improving his pedagogy. The third teacher was a young and highly skilled practitioner with a thirst for learning. His English teaching was genre-based and he was looking for ways of extending boys' higher order thinking. The fourth teacher was an experienced teacher of senior English who had, exclusively, adopted a critical literacy approach. He was keen to see whether the researcher's literacy and thinking tools, and associated assessment protocols might "fit in" with his existing approach.

The second characteristic common to all volunteers was the desire to change. Among the schools generally, and more acutely within the volunteer group, there was a ground swell of opinion that their pedagogy and assessment protocols that had been used for many years, needed to change. Equally, there was little consensus or rationale for how these changes might occur.

Vignettes

The following brief vignettes illustrate how these middle and secondary school teachers used literacy and thinking tools, and how they used these tools as assessments items. These vignettes are followed by thematic analyses based on observations and interview that link to the vignettes.

Teacher A: An assessment item linked to report writing

Understanding text structure at the paragraph and whole text level can be problematic for many students because the texts they encounter, especially in the secondary school, often differ markedly from the familiar narrative and oral discourse structures of junior and middle schools. Developing an understanding of text structure is crucial, as this understanding affects readers' recall and comprehension (Kintsch, 1994).

Teacher A, a middle-school teacher, used a *Brainstorm* tool (Whitehead, 2001) and a *report draft writing frame* (Whitehead, 2008) to assist her middle-school students understand the structure of report texts, and to help them write these texts. The *Brainstorm* tool assisted students collect, group and label information prior to writing, and the draft writing frame provided a scaffold that corresponded to the structural conventions of report texts. Specifically, the *report draft writing frame* signalled to students that they might first draft an *engagement* section designed to "hook" the reader, followed by a subject *classification* section, and a *body,* that included topical paragraphs. These paragraphs were to include topic, detail and example components. The teacher introduced the *Brainstorm* tool and *report draft writing frame* in the context of a series of instructional reading lessons that employed well-structured report texts. Consistent with the key principle underpinning the definition of ecologically valid assessments, an end of topic assessment item asked her students to use a *Brainstorm* tool prior to writing a topic-linked report.

An analysis of the topic-test scripts indicated that her students were able to use a *Brainstorm* tool, and write reports that included an *engagement* section with an identifiable, first-person, writer's voice. The analysis also revealed that most students were able to use the present tense in the body of their report, and construct topically structured paragraphs. After the test, students reported feeling confident about themselves as writers because the *Brainstorm* tool and their knowledge of text structure had provided them with some surety in terms of content and structure. This was especially the case for less able/reluctant writers, as illustrated in the following transcript from a student who with others was asked to write a report about reports (!) at the beginning of the class focus on report writing.

> *Reports are dumb. They are so hard to write and who wants to know what a 13 year old girl has to say. Anyway, I don"t like writing. But in some way reports are good. They tell people that we have won the rugby game or something. Finally reports are good to read, but not to write.* (Complete script)

In addition to conveying her negative attitude, the student illustrates her confusion between recounts and reports. Both these findings informed the teacher and influenced her planning and teaching. For her assessment-test script, this same student chose to write a report about cats, beginning:

> *What is a cat? Well the cat you might be thinking of is a harmless little pussy that runs around your bedroom trying to catch a snack. But that's not what I'm talking about. There are bigger cats. Want to know about cats? Well read on and be amazed!* (Followed by six well-structured paragraphs).

Teacher A illustrated how literacy and thinking tools can be incorporated in the design of ecologically valid end-of-topic tests. This item provided a measure of how well her students could use a *Brainstorm*, a *report draft writing tool*, and how well they could write reports.

Teacher B: Assessing vocabulary

Corson (1999) notes that "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). Texts can

be challenging because they differ greatly with respect to contextual support for unfamiliar vocabulary. Texts are especially problematic when there is an inequality between home and school vocabulary, both in terms of the status and acquisition of academic vocabulary. Given Olson's (1997) contention that oral participation is a key to vocabulary growth, and given that vocabulary provides the conceptual links for learning, it was unsurprising that Teacher B used a *Matching Task* tool with a Year 5 class to enhance her students" vocabulary. This tool assisted her students to:

- Acquire vocabulary that described the characteristics of narrative participants ("caring", "kind", "thoughtful", "greedy", "mean" and so on);
- Substantiate these participant characteristics from a text;
- Identify text themes and messages;
- Write a character description, with the help of a *character description draft writing frame* (Whitehead, 2008).

Consistent with the key principle underpinning the definition of an ecologically valid test, she used the *Matching Task* tool during a series of lessons, and as an assessment item (see Figure 1).

Instructions: Read the attached short story. As you read, think about the characteristics of the main characters. Match the words in the left column with events from the story involving these characters in the right column, and add the name of the character. Write your answers in the ANSWERS box below.

Characteristic	**Events**
1. caring	A. He took only one necklace
2. kind	B. She cooked more rice balls for her husband after the first ones rolled away.
3. mean	C. He took all the jewels
4. greedy	D. He tricked the mouse

ANSWERS			
1	*2*	*3*	*4*

Example: 1. C. (John Brown)
Attached short story.

Figure 1. *Matching Task* item

Again, this teacher reconstructed how her students learnt by using an ecologically valid test item in the end-of-topic test.

Teacher C: Using a Factual Meaning Grid

Two Year 9 and 10 English teachers designed ecologically valid pre-test and post-test items based on their use of a *Factual Meaning Grid* tool (see Figure 2). This tool was used to assist students gather and process information while reading a report text. The students had used this tool only once before encountering it, again, in the pre-test. Consistent with the key principle underpinning the definition of an ecologically valid test item, the teachers used a *Factual Meaning Grid* as a teaching tool (once again) and asked students to complete and use

another grid as an end-of-topic test item (see Figure 2) based on a supplied report text. Unfortunately, students had difficulty completing this test item, perhaps because it had been co-constructed with the teacher only twice before. However, test results indicated that students had mastered the use of the *Factual Meaning Grid* as a reading (note-making) and information-processing tool.

Instructions: As you read the report text about whales, write the types of whales across the top of the Factual Meaning Grid, and some features of those whales down the side. Then complete the grid with √, X or •, and write statements beside and below your completed grid that show you have thought about what you recorded on the grid.

Features of the whales	Type of whale	Type of whale	Type of whale	Write "All", "Some" or "None" statements in the spaces below

Write compare and contrast statements in the space below.

Report text about whales.

Figure 2. *Factual Meaning Grid* item

Teacher D: Using a Concept Frame

A Year 10 English teacher used a *Concept Frame* literacy and thinking tool (see Figure 3) during a series of lessons, to help his students define character-types in the myth genre. Initially, the *Concept Frame* was co-constructed with his students during these lessons. Later, students used the tool as an independent learning tool (as they would require doing during the end-of-topic test). The inclusion of an assessment item that required students to complete a *Concept Frame* was foreshadowed.

Again, the tool used in this test item had been used in class as part of the teaching/learning process, and its inclusion in the end-of-topic test had been foreshadowed.

Teacher E: Using a Timeline

This Year 5 classroom teacher used folktales to help her students understand simple narrative text structure, and to further develop their ability to construct causal inferences. She began by reading a simple folktale (*Little Red Riding Hood*) with her students and completing a simple *Plotline* (completing only the top and bottom boxes illustrated in Figure 4). Students then

Instructions: Define the meaning of a hero / villain as accurately as you can by completing the following Concept Frame. Then use the completed Concept Frame to help you write your definition of a hero. Make sure your definition uses information from the "examples" you have listed.

A hero is (is a..)…. *1* *2* *3*	*A hero can…..* *1* *2* *3*
Examples of different kinds of heroes are… *1* *2* *3*	*A hero has (has a…) ….* *1* *2* *3*

My definition of a hero

Figure 3. *Concept Frame* item

Chose their own folktale and completed a simple *Plotline* independently. Consistent with the key principle underpinning the definition of an ecologically valid test, the students completed a simple *Plotline* cloze assessment item in a mid-topic test.

This simple *Plotline* which engaged her Year 5 students in procedural thinking (episodes/events over time) did not present them with any great cognitive challenge. In contrast, complex *Plotlines* are more challenging because they differentially engage explanation (psychological) rather than simply procedural thinking. Inferencing is fundamental to the comprehension of explanations and to the completion of a complex *Plotline*. So, after further work with the researcher this teacher returned to the *Little Red Riding Hood* text, selected an episode (see Episode 2 in the bold box in Figure 4) and helped students construct causal inferences using a problem, response, action, outcome-rubric (see centre section in Figure 4). The interesting aspect of this analysis was the use of "because", which prompted students to compose the causal inferences. The topic, test item associated with these lessons included a blank complex *Plotline* (that is, the time, setting, episode and resolution boxes, and the problem, response, action and outcome words in bold), a new folktale, and instructions to complete the tool.

Again, this teacher tested as she had taught. She gave her students an opportunity to select an episode and demonstrate their ability to engage in causal thinking through a complex *Plotline* that she had used as a teaching tool. She tested their knowledge of narrative structure, their ability to inference and their ability to use a complex *Plotline,* all in the one ecologically valid test item.

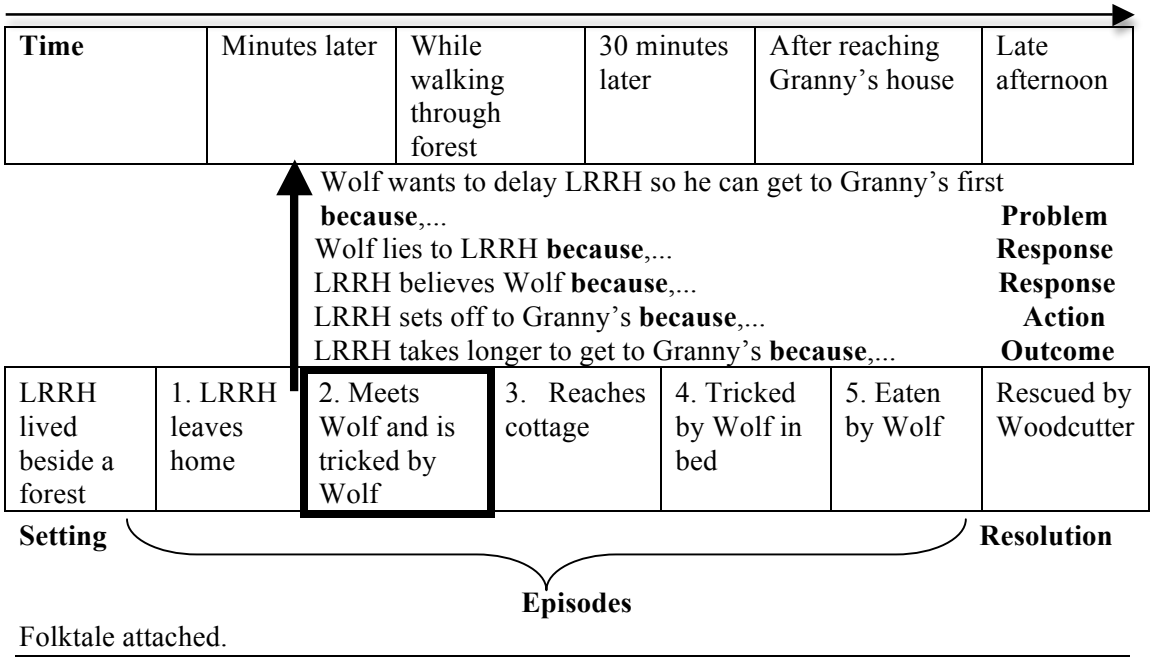

Figure 4. Complex *Plotline* item

A THEMATIC ANALYSIS OF OBSERVATIONS AND INTERVIEWS

A thematic analysis of observation and interview data associated with these brief vignettes was conducted to explore the impact of the praxis research on teachers and students. Five key themes emerged from this analysis. These are affective challenges, cognitive challenges, scaffolding, structural changes and the re-definition of knowledge.

Theme #1: Affective challenges

Teachers are unlikely to adopt new pedagogy, or change their assessment protocols, unless they feel it is worth the risk, and unless they are willing to accept the emotional challenges associated with change. Together, these pre-requisite dispositions speak of teachers' attitudes toward the research and the researcher. In a general sense, the use of the literacy and thinking tools, and the construction and implementation of ecologically valid test items were accompanied by positive responses. For example, two Year 9 and 10 English teachers, who co-constructed and administered a test containing ecologically valid test items, decided to use a similar type of test as part of their next junior writing unit, and to extend the use of the tools and ecologically valid tests into the secondary school. A third teacher reported that the way he taught the topic, which had been modified as a result of using tools and re-designing the assessment was:

> …going really, really good… it's a hell of a lot easier to do this (than last year), it's more structured, the students like it, you don't go off on tangents, [and] the resourcing of it is a lot easier – journals etc. using them in a constructive manner.

A fourth teacher noted:

> It's a bit more than teacher-made tests; what we've done is taken teacher [tools] ideas and packed them in a unit and delivered them in an ecologically sound pre-test to post-test module.

Students also responded positively to the use of tools as teaching and assessment devices. One student quipped, "That's the best lesson you've taught, sir," and another, "…that was better than your usual lesson."

Theme #2: Cognitive challenges

There are also cognitive challenges associated with using literacy and thinking tools, and tool-linked test items for the first time. These challenges were illustrated by Teacher E, who initially chose to use a simple *Plotline* tool, and to design a cloze assessment item that aligned with the use of that tool. The teacher remarked that the simple version was "all she could handle" right now. This was despite the researcher modelling the complex *Plotline* that aligned with the teacher's learning intention of improving students" inferential thinking and the abilities of his students.

The research also indicated that there are cognitive challenges involved in moving from procedural thinking (simply recording episodes over time as prompted by the simple *Plotline* tool), to the use of tools that engage "higher-order" causal thinking (explaining how and why characters behave as they do, as prompted by the complex *Plotline* tool). From a pedagogical perspective, preparing students to answer assessment items that differentially engage causal thinking requires teachers to first engage them in dialogues designed to explore characters" motivational goals, involuntary thoughts and feelings and physical actions (Warren et al., 1979). Teacher E's initial use of a simple *Plotline* tool to record time and episodes aligned well with her default transmission approach because, cognitively, it involves little more than "lower-order" recall thinking. But, as a result of her involvement in the action research process, and her later use of a complex *Plotline* tool as a means of helping students explain characters" behaviour, she was able to involve her students in co-constructing inferential meanings. This was, cognitively, a more challenging task. What this "cognitive challenge" theme suggests is that the uptake of ecologically valid assessment measures is predicated upon successfully mastering a range of cognitive challenges.

Theme #3: Scaffolding

What made these affective and cognitive challenges all the more demanding was the undue haste with which some teachers of English introduced tools, and tool-linked forms of testing. They failed to adequately scaffold their students" learning. This "scaffolding" theme emerged, especially from an analysis of interview and observation information obtained from

Teacher C, who used a tool twice only before his students were asked to use it independently in the context of an assessment item.

Typically, the time between teachers introducing literacy and thinking tools, and their use in test items, was too compressed. This compression reflected a pedagogical history of transmission "to cover the content", and a failure to use tools as a means of developing independent, literate thinkers. Consequently, when the tool was used in Teacher C's class as a test item, students were unsure how to proceed. Teachers became aware of the problems associated with introducing tools with undue haste through reflection. During interviews, one teacher indicated that he had taught some of the tools "…poorly, too quickly and with not enough deliberateness". He linked students' below-average performance on some assessment items to "poorly taught tools".

The use of literacy and thinking tools as assessment items presupposes that the teacher-to-student transfer of expertise associated with their use has occurred. Discussion with teachers around this theme suggests they might benefit from further theoretical understandings about the use of tools first as "teaching-focused" and later as "learner-focused", and the importance of scaffolding this transition.

Rather than seen as a negative aspect of the research, teacher reflection about links between how a topic was taught and how a topic was tested, together with an item analysis of test results, seemed to strengthen their pedagogical practice. They began to appreciate that the undue haste with which they had introduced tools had prevented them from scaffolding students from dependent to independent behaviours. What emerges from an examination of this theme is that associated with the use of ecologically valid test items is the need for meta-cognitively aware students and patient teachers.

Theme #4: Structural changes

A major theme to emerge from an analysis of the observation and interview data was how schools" structural characteristics impacted on the use of literacy and thinking tools and associated, ecologically valid test items. In schools with "transmission teaching" cultures, the use of literacy and thinking tools, and especially their inclusion as test items, raised challenges for teachers. As one teacher noted:

> As a new teacher still coming to grips with the culture of learning at this school - and with the way English is currently taught in the school, I did find it difficult to introduce some of the things [tools]…. The students have developed a culture of demanding to be spoon fed information and so attempting them is often difficult. And the exams for this year are the same ones as used last year.

Although the research assisted teachers to reflect on their practice, the culture of the schools, and the value they placed on national high-stakes testing, proved a resistor to change.

Theme #5: The redefinition of knowledge

A final theme to emerge was the effect of literacy and thinking tools on redefining knowledge. To this extent the introduction of tool-linked test items worked against the

researcher's acceptance and the inclusion of tools as test items. However, as the volunteers changed from transmission to co-construction approaches through the use of literacy and thinking tools, they gained in confidence and began to treat knowledge as a verb, both in their teaching and assessment practices. Consequently, their new assessments required students to work with what they knew rather than recall what they knew, which had been the *force majeure* of the past. The analysis of interview data suggested that the restructuring of historic tests and test protocols is not only about designing ecologically valid items, it is about accepting a new paradigm for education.

CONCLUSION

New Zealand secondary schools have a history of administering topic-focused, summative assessments, in English and other subjects, that seldom reflect the pedagogy used to help students come to know a topic, and use their knowledge of that topic. Results from these tests rarely seem to influence what teachers (re)teach, or how they teach. This article has explored the efforts of teachers who, in the context of high-stakes, international and national summative assessment protocols, designed and administered local summative tests that incorporated ecologically valid test items.

The design of their test items, generated as part of a praxis action research study, reflected literacy and thinking tools they had used to teach their topics. The items acknowledged process as an integral and valid component of teaching, and assessment. The research shows that ecologically valid test items can be included in the re-design of historic topic tests that allow teachers to measure subject-specific content, subject-specific knowledge about text structures, and types of subject-specific thought processes. Although the use of these tools and associated test items was risky to those involved in the research in both a personal and pedagogical sense, and risky in respect to the potential impact on high-stakes test results, these teachers deemed it a risk worth taking.

REFERENCES

Amrein, A. & Berliner, D. (2002). High-stakes testing, uncertainty and student learning. *Education, Policy Analysis Archives, 10*(8). Retrieved June 10, 2008, from www.http://epaa.edu/epaa/v10n18/.

Brock-Utne, B. (1996). Reliability and validity in qualitative research within education in Africa. *International Review of Education, 42*(6), 605–621.

Castells, M. (2000). *The rise of the network society* (2nd ed.). Oxford: Blackwell.

Chow, S. (1987). Science, ecological validity and experimentation. *Journal for the Theory of Social Behaviour, 17*(2), 1 – 6.

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.

Cizek, G. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice, 20*(4), 19-27.

Corson, D. (1999). *Language policy in school.* Mahwah, NJ: Lawrence Erlbaum.

DeWalt, K. & DeWalt, B. (2002). *Participant observation: A guide for fieldworkers.* Walnut Creek, CA: AltaMira Press.

Freebody, P., & Wyatt-Smith, C. (2004). The assessment of literacy: Working the zone between "system" and "site" validity. *Journal of Educational Enquiry, 5*(2), 30–49.

Gambell, T. (2004). Teachers working around large-scale assessment: Reconstructing professionalism and professional development. *English Teaching: Practice and Critique, 3*(2), 48-73.

Gibson, J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Haber, R. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information processing. *Behavioral and Brain Sciences, 6*, 1-54.

Higginson, C., Arnett P. & Voss W. (2000). The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Archives of Clinical Neuropsychology, 15*(3), 185–204.

Kemmis, S. (1988). Action research in retrospect and prospect. In S. Kemmis & R. McTaggart (Eds.), *The action research reader* (3rd ed.) (pp. 27-39). Geelong, Victoria: Deakin University Press.

Kintsch, W. (1994). The psychology of discourse processing. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 721-739). New York: Academic Press.

International Association for the Evaluation of Educational Achievement (2006). The Progress in International Reading Literacy Study (PIRLS). Washington: US Department of Education.

Lather, P. (1986). Research as praxis. *Harvard Educational Review, 56*(3), 257-277.

Lyotard, J. (1984). *The postmodern condition: A report on knowledge*. Manchester: Manchester University Press.

Ministry of Education & University of Auckland (2003). *Assessment strategies for teaching and learning: He punaha aromatawai mo te whakako me te ako. Version 2.* Wellington: Learning Media and the Learning Media Trust of New Zealand.

Neisser, U. (1976). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Olson, D. (1997). Talking about text and the culture of literacy. In B. Davies & D. Corson (Eds.), *Oral discourse and education* (pp. 2–9). Boston: Kluwer.

Organization for Economic Cooperation and Development (2006). Program for International Student Assessment (PISA). Washington: Organization for Economic Cooperation and Development.

Schensul, S., Schensul, J., & LeCompte, M. (1999). *Essential ethnographic methods: Observations, interviews, and questionnaires* (Book 2 in Ethnographer's Toolkit). Walnut Creek, CA: AltaMira Press.

Warren, W., Nicholas, D., & Trabasso, T. (1979). Event chains and inferences in understanding narratives. In R. Freddle (Ed.), *New directions in discourse processing* (Vol 2) (pp. 23-52). Hillsdale, NJ: Lawrence Erlbaum.

Whitehead, D. (2001). *Top tools for literacy and learning*. Auckland: Pearson Education.

Whitehead, D. (2004). *Top tools for teaching thinking*. Auckland: 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 Aoteraroa* 57, 3–12.

Whitehead, D. (2008). *Top tools for literacy and thinking*. Auckland: Pearson.

Wright, N., May, S., Whitehead, D., Smyth, J., & Smyth, S. (2005). *Secondary Schools" Literacy Initiative Research Evaluation: Sustaining literacy initiatives in secondary*. Hamilton: Wilf Malcolm Institute for Educational research, University of Waikato.

Manuscript received: August 2, 2008
Revision received: November 20, 2008
Accepted: December 5, 2008