

Assessment: what are the cultural issues in relation to Pasifika, Asian, ESOL, immigrant and refugee learners?

Stephen May
University of Waikato

A key challenge for those connected to the assessment enterprise is to ensure the accuracy of information and the appropriateness of decisions for *all* groups within a given society. This is particularly important for cultural and linguistic minority groups, who have too often been poorly served, positioned, and subsequently represented in deficit terms by standardised educational assessment measures.

Evidence internationally suggests that minority groups tend to be consistently over-represented at a national level in the lowest levels of educational achievement when compared to other groups. Where assessment measures are used as a basis for streaming in individual schools, such groups are more often to be found in lower 'ability' classes. They are also more likely to be directed to special education services.

These patterns are broadly replicated in New Zealand, particularly in relation to Māori, Pasifika, refugee, English Language Learners (ELLs) and some Asian students. This paper will not address directly issues to do with Māori learners, as this is the focus of the paper by Bishop and Mahuika. However, it will explore the implications of assessment for the remaining minority groups, with a particular focus on the interrelationships between culture, language, and assessment practices. It is expected that issues identified in this paper will provide a basis for a more inclusive and equitable approach to the assessment of such students over the medium to long term in New Zealand schools.

In order to address these issues, the paper will be divided into the following sections:

1. The cultural locatedness of assessment

Drawing on international research, this section will explore the degree to which majoritarian cultural norms, and related biases, still underpin various forms of educational assessment, advantaging students from majority populations and disadvantaging students from minority populations.

2. The language of assessment

Drawing on international research, this section will explore the role of language(s) with respect to assessment. Assessment practices are still predominantly monolingual and/or conducted in the majority language of a given society. As such, assessment results for minority, bilingual and/or ELL students may not be an accurate reflection of their academic ability but are rather confounded by 'the language of the test'.

3. Assessment and minority groups in New Zealand

This section will summarise the still relatively limited research undertaken in New Zealand on the differential effects of assessment for Pasifika, ELL, refugee and Asian students in light of the wider issues identified in the two previous sections.

4. Implications for policy and practice

The final section will highlight the implications for assessment policy and practice in New Zealand in light of the above, as well as identifying current gaps in research and future research possibilities.

1. The cultural locatedness of assessment

What is striking in the international literature on educational assessment is how much of its focus is directed (or limited) to issues of technology and form – that is, to the technical aspects of its development and implementation. As Delandshere (2001) observes, those involved in educational assessment seldom question or explore the role that assessment plays in the larger society. Current assessment practices are debated with regard to the impact they have on defining or narrowing the curriculum, for example, or on the need for different forms of assessment – summative and formative assessment practices, most notably. But only in rare cases have scholars examined the functions that assessment serves in validating and reproducing certain forms of knowledge or ideas at the expense of others.

This tendency toward a ‘technicist’ approach to educational assessment, disconnected from wider social and cultural contexts and considerations, arises from the normative assumptions that underpinned its development in the nineteenth century. As Broadfoot and Black (2004) note, these key ‘objectivist’ principles included the following:

1. That it is right, ‘objectively’ to seek to identify relative levels of student performance as the basis for educational selection.
2. That it is possible to undertake such identification with a sufficient degree of ‘objectivity’ that it provides a broadly fair outcome for the candidates affected.
3. That the quality of such assessment is embodied in notions of reliability and validity.
4. That students’ scores on national examinations and tests provide a valid indicator of the quality of institutional performance.
5. That it is possible usefully to compare the ‘productivity’ of individual education systems through international comparisons.

So taken for granted are these underlying assumptions, Broadfoot and Black observe, that they are rarely, if ever, questioned, let alone critiqued by those involved in educational assessment. And yet, these technicist, objectivist, notions often fail to address adequately, and often specifically mask, the inherent cultural locatedness of educational assessment practices and their uneven application and effects across different cultural groups.

The response has been the development over the last decade of a still nascent, but nonetheless important sociocultural critique of educational assessment. Much of this sociocultural critique is focused on standardized, summative forms of assessment – that is, assessment *of* learning. Critics argue, as we shall see, that summative assessment can have significant differential and often deleterious effects on ethnic,

cultural and linguistic minority groups, precisely because they reflect, reproduce and privilege dominant cultural understandings and norms.

Formative assessment, or assessment *for* learning, is often seen as a potential rejoinder to summative practices in this respect, given its focus on enhancing student development within classrooms and its affinity with good teaching practices (Crooks, 2002). However, formative assessment – widely practised, as we know, in New Zealand schools – does not escape charges of cultural normativity either.

The following areas/issues have been identified by sociocultural critiques of educational assessment in the international literature as particularly significant.

Knowledge as normative, bounded and static

Drawing on the work of Bourdieu, Bernstein, and Foucault, amongst others, sociocultural critiques of assessment highlight the static, universalist and unitary conceptions of knowledge that underlie much educational assessment – the assumption, in effect, that all social and cultural groups have the same access to the particular forms of knowledge being assessed. Such unitary conceptions of knowledge, critics argue, underplay or ignore the differential nature of knowledge within and across cultural groups, as well as the social and cultural hierarchies in the wider society which mediate and/or limit access to such knowledge, depending on the social and cultural location of the learner (see, for example, Gipps, 1999; Delandshere, 2001; Broadfoot and Black, 2004). The result, they argue, is that minority groups do not fare as well in education as they might. As Delandshere (2001: 130) summarizes it, from a US perspective:

As long as school teaching and learning are subordinated to such dubious assessment practices, it will yield forms of teaching and learning that are narrow, monolithic, and unmotivating for most teachers and students. This state of affairs can only disadvantage those students who do not already have access to other ways of learning or, in Bourdieu's terms, to cultural and economic capital, hence playing out and reproducing current socio-economic structures. After decades of such practices the representations of teaching and public school learning, by both teachers and the public, have been, for the most part, fossilised. These representations assume a body of knowledge in different areas to be acquired by students and see teachers as implementing a set of activities to ensure that this knowledge is acquired in the forms that are represented on the tests. The tests confirm and reify the importance of that knowledge.

This critique raises, in turn, a range of key questions with respect to adopting a more inclusive, heterogeneous, and, by extension, equitable view of knowledge as it underpins curricula and related educational assessment. Paraphrasing Gipps (1999: 365), these can be summarized as follows:

| Curricular Questions | Assessment Questions |
|---|--|
| Whose knowledge is taught? | What knowledge is assessed and equated with achievement? |
| Why is it taught in a particular way to this particular group? | Are the form, content, and mode of assessment appropriate for different groups and individuals? |
| How do we enable the histories and cultures of minorities to be taught in culturally responsible and responsive ways? | Is the range of cultural knowledge reflected in definitions of achievement? How does cultural knowledge mediate individuals' responses to assessments in ways that alter the construct being assessed? |

To these questions can also be added the question of power. As Hawk et al. (2001: 4-5) summarize it, students who originate from 'dominated' societal groups can be either 'empowered' or 'disabled' as a direct result of their interactions with teachers and in relation to educational assessment. The longstanding, ongoing over-representation of Māori, Pasifika, ESOL and/or refugee children in the lower echelons of educational achievement in New Zealand – most notably, in the so-called 'literacy tail' – suggest that such 'dominated' groups remain educationally disadvantaged as a result of teaching and assessment practices in New Zealand, *despite* key changes to assessment practices over the last decade.

A key issue that emerges here is the discourse of 'locus of control', one that stresses the importance of students being supported and encouraged to take responsibility for being actively involved in their own learning and assessment. Locus of control is most evident in more formative and classroom-based assessments (see below) and is closely linked to self-efficacy, motivation and educational success. However, as Hawk et al. observe, it is still more difficult to establish in education settings where students are from low socioeconomic areas and minority cultures. New Zealand appears to be no exception in this regard.

Culture and equity

The ongoing disadvantage for subordinated or dominated societal groups in New Zealand, as elsewhere, highlights the importance and role of culture in relation to educational assessment policy and practice. As Gopaul-McNicol and Armour-Thomas (2002: 6-7) summarize it, culture is a socially constructed phenomenon that enables an understanding of the way of life of any social group. It is a multifaceted concept, the defining characteristics of which are many. Key dimensions of culture include here:

- The judgmental or normative dimension (values, standards of social group)

- The cognitive dimension
- The affective dimension
- The skills dimension
- The technological dimension

Culture influences the ways in which people construct knowledge and create meaning from experience (how they think about things, reason, and solve problems), and this in turn relates directly to the ways in which individuals learn and teach in both informal and school settings. As Solono-Flores and Nelson-Barber (2001) argue, however, while current approaches to handling cultural diversity aim to ensure equitable testing, they fail to address the fact that culture shapes the mind. For example, estimating systematic score differences across cultural groups focuses on correcting for item bias and ensuring assessment comparability across cultures. However, the *process* of assessment development is overlooked by these approaches. Rather, the finished version of a test originally developed for a particular sector of the population is most often simply adapted for a different population or translated into another language. The cultural locatedness or specificity of the test, and the processes of its construction, remain unaddressed.

The challenge for equity in assessment is thus to ensure that the judgments made about the behaviour of individuals and groups are accurate and that the decisions made on the basis of assessments do not intentionally or unintentionally favour one cultural group over another. However, this proposition may be difficult to operationalize, for there are many culturally related factors – particularly relevant to the standardized assessment process – that could invalidate the information for some cultural groups. As Gopaul- McNicol and Armour-Thomas (2002: 11-12) outline, these include the following:

- *Subject sampling*: a common practice in test construction is to include an adequate representation of subjects in the norming process. However, what constitutes adequate representation of some cultural groups remains problematic (e.g. social class differences, or the confounding of ethnicity and/or language background with culture)
- *Behaviour sampling*: the idea that tests measure what their designers claim it measures, via a representative sample of items. However, cultural groups may have conceptions or understandings of a construct that differ from those represented in assessment measures. Ensuring a representative sample of items to measure a particular construct thus becomes meaningless for those children for whom the conceptualization of the construct is invalid.

This in turn raises the idea of the *cultural equivalence* (or fairness) underpinning assessment tasks (Okazaki & Sue, 1995). Gopaul-McNicol and Armour-Thomas argue that cultural equivalence may be undermined by the following factors:

- *Contextual non-equivalence*: whether the sociolinguistic factors such as norms of discourse are not familiar to all children tested
- *Conceptual non-equivalence*: the extent to which task attributes do not have the same meaning for children from different cultural backgrounds
- *Linguistic non-equivalence*: the extent to which the language used in standardized assessments does not have the same meaning for children from different cultural backgrounds.

Gopaul-McNicol and Armour-Thomas conclude by highlighting two further issues of possible inequity in relation to the notion of cultural equivalence and assessment practices:

- When some cultural groups have greater opportunities to practise tasks similar to those used on assessment measures
- When opportunities for exposure to high-quality teaching and curriculum are differentially allocated.

Minority students are often disadvantaged in relation to both factors, especially the latter. This is because a disproportionate number of students from minority backgrounds, *on the basis of educational assessments*, are allocated to 'lower ability' classes. This has significant implications for the long-term educational achievement of such students since, as research from the US highlights, in lower ability classes:

- Teachers fragment topics into isolated bits of information, diminishing overall coherence of the lesson
- Assignments often require rote memorization and little higher order skills such as critical thinking, abstract reasoning and problem solving
- More time spent on management than instruction
- Curricula focused on low-level skills and less rigorous topics (see Gopaul-McNicol & Armour-Thomas, 2002: 77 for further discussion).

New Zealand does not escape this prejudicial nexus of teaching and assessment for minority students since, as we shall see in Section 3, the use of assessment measures for streaming purposes remains widespread, particularly in secondary schools, as is the related over-representation of Māori, Pasifika, ESOL and refugee students in lower band classes.

Cultural validity

The lack of cultural equivalence in assessments and the related differential in student achievement and placement by ethnicity raises the wider issue of the cultural validity of assessments for ethnic minority groups. Discussing the notion of

cultural validity in relation to science assessments, Solono-Flores and Nelson-Barbour (2001: 555) describe it as follows:

By cultural validity we refer to the effectiveness with which science assessment addresses the sociocultural influences that shape student thinking and the ways in which students make sense of science items and respond to them. These sociocultural influences include the sets of values, beliefs, experiences, communication patterns, teaching and learning styles, and epistemologies inherent in the students' cultural backgrounds, and the socioeconomic conditions prevailing in their cultural groups.

Solono-Flores and Nelson-Barbour argue that while a cognitive perspective on assessment can provide valuable information about the students' use of content knowledge and problem-solving strategies, it cannot provide information on the students' epistemologies and the sociocultural factors that influence how students construct knowledge. Thus addressing students' epistemologies is crucially important here. This also relates to the recognition of different cultural worldviews and students' life contexts and values, a feature that has come to be associated with culturally responsive pedagogies.

So-called culturally responsive pedagogy is based on developing educational methods that are situated in students' cultural experiences. Such an approach is especially relevant in the case of cultural groups for whom knowledge is highly contextualized, as in the case of many indigenous peoples (including Māori). From this, assessments developed for a specific cultural group should be sensitive to its ways of knowing and traditional knowledge. However, how accurately assessment developers understand the ways of knowing and the traditional knowledge becomes critical to attaining cultural validity. Stereotypes and unrealistic expectations about the students' knowledge and skills can result from the simplistic assumption that all the individuals from a culture are familiar with certain knowledge associated with it. We see this, for example, in the common stereotypical construction of Māori and Pasifika students as primarily 'kinesthetic' in their approach to learning.

In conclusion, Solono-Flores and Nelson-Barbour argue that cultural validity goes beyond the concept of fairness in testing as a criterion for test validity. This is because correcting for cultural bias, promoting the participation of ethnic minorities in pilot student samples, and providing accommodations for linguistic minorities are in the end simply remedial strategies that address cultural differences not considered in an assessment's original plan. Ideally, if cultural validity issues were addressed properly at the inception of an assessment and throughout its entire process of development, they argue, there would be no cultural bias and providing accommodations for cultural minorities would not be necessary. Cultural validity thus establishes the need for incorporating the reasonings inherent to a sociocultural perspective in assessment practices (2001: 556-557).

What are the implications of developing and applying a culturally valid approach to various assessment practices? The final part of this section explores these implications in relation to standards-based, formative, and alternative (classroom-based) assessments.

Standards-based assessment

Standards assessment, particularly high stakes testing, is consistently the most criticized form of assessment with respect to issues of cultural validity. Following on from Solono-Flores and Nelson-Barbour's observations, a key issue identified is the process by which standards are developed in the first instance, given that such standards must reflect some form of consensus. Sociocultural critique highlights that this consensus is a normative one, invariably (de)limited to the majority group. As Moss and Schutz (2001: 38) observe:

In the most basic sense, standards-based assessments are supposed to derive their warrant from a community consensus. It is fundamentally because the community has agreed on a set of standards that it is reasonable to use them to orient assessment instruments. Yet, we do not really know much about the actual process of achieving such a consensus. We do not know what sort of agreement is reasonable to expect or what the implications of this kind of agreement might be for justifying the interpretation, use, and ultimate impact of standards-based assessment.

From this, Moss and Schutz ask the following key questions:

- To what extent are such standards capable of supporting such decisions?
- To what extent are assessment developers and, ultimately, judges forced to exceed the level of agreement that was reached in the standards creation process in order to make determinate statements about individual performances?
- If all the participants in the standards creation process examined the concrete decisions (and it is difficult to know who would qualify as a participant), would they consider them consistent with the standards they had developed?

A key target for explicit criticism in this respect is *No Child Left Behind* (NCLB). Critics of NCLB have asserted that its focus on student testing reflects and reinforces a narrow, normative conception of the curriculum, and related curricular knowledge. As a result, nonacademic areas such as sports, music, and art have become marginalized or eliminated, while the focus on testing and accountability actively harms students who learn differently, including culturally and linguistically diverse learners. This would appear to be confirmed by numerous studies (cited in Fusarelli, 2004) that have found that high-stakes testing limits the flexibility of teachers and decreases the creativity of teachers and students by emphasizing drill-and-kill skills sets, particularly for those students in low stream classes (who, as we have seen, include an over-representation of minority groups).

Formative and classroom based assessment

Formative assessment, and related criterion-based assessment, is often touted in the literature as more effective in highlighting what students actually know – what they bring to the learning context (Black and Wiliam, 1998; Black, 2000). Certainly, this rationale has been prominent in the development of curriculum-based assessments in New Zealand over the last decade, most notably in relation to NCEA (Crooks, 2002). The key principles of curriculum-based assessment, as summarized by Black and Wiliam (1998: 44) are:

- Assessment should reflect key learning aims and identify key learning needs
- The main purpose of assessment is formative
- The main focus is the individual learner's individual attainment in relation to the criteria set, but can also be informed by norm data of other learners working on the same curriculum
- Assessment should be frequent to track progress over time – the gradient of learner success is the key indicator.

From a related angle, Gopaul-McNicol and Armour-Thomas (2002: 82) argue that this form of 'curriculum-embedded' assessment measures student performance on the very tasks used for instruction. Four key principles are involved for them:

- Assessment procedures are designed to mirror the nature of current instruction
- Assessment tasks are drawn from the instructional materials
- Results assist teachers to remediate student difficulties on materials previously covered
- Assessment data helps the teacher to design more effective instructional strategies for subsequent coverage of instructional content.

With its emphases on individual learning of specific domains of knowledge within the classroom context, and its allied feedback and feedforward dimensions, curriculum-based assessment is also often heralded as a more equitable and inclusive assessment process for students, especially minority students. Of particular significance here are so-called 'alternative', 'authentic' or 'performance' assessment practices, including portfolio assessment. Such assessment involves a collection over time of multiple performances to provide evidence of growth and learning. It requires learners to: actively shape the assessment process by selecting which of their performances will be evaluated; collaborate with other students and the teacher in identifying criteria for evaluation; engage in self- and peer assessment; and reflect on an ongoing basis on their learning (Fox, 2008).

The assumption here is that students' (diverse) backgrounds, their learning styles, and their cultural preferences for different ways of participating in classroom activities can be better recognized and included in these more informal, negotiated, assessment processes. It is also purported that the gap in academic achievement in

relation to ethnicity narrows when such measures are implemented (Gopaul-McNicol & Armour-Thomas, 2002; 82; see also Shepard, 2000). This is because students can better achieve if assessments allow them to use those modes of expression in which they are most fluent, and this means that multimodal expressions, involving varied combinations of art, craft, oral and verbal modes, should be recognized in assessment practices (Broadfoot & Black, 2004).

However, despite the potential benefits of alternative forms of assessment, Black and Wiliam 1998 (see also Broadfoot and Black, 2004) highlight that there remains little research evidence, beyond teacher report, that establishes clear learning advantages for students. There is also the concern that such approaches require levels of expertise that may not be fully developed in practice. As Fox (2008: 104) summarizes it, with clear echoes of ongoing criticism of NCEA in the New Zealand context:

Students may not have sufficient experience with alternative assessment to value 'authentic' learning, and may instead subvert the intended purposes of an assessment..., living up to what they perceive are external expectations to perform for marks – to 'simulate' reflection, thinking, learning, rather than to actually reflect, think or learn. Teachers may neither be able to connect ongoing assessment in support of individual learning with curricular goals, nor to effectively communicate information about development and growth to parents and other stakeholders.

With respect to how effectively teachers carry out formative assessment, Leung and Lewcowicz (2008: 311-312) argue that the following key questions need to be addressed:

What do teachers do when they carry out formative assessment?

- Do teachers do something different from, or in addition to, their everyday teaching when they assess formatively? (If formative assessment is really the same as 'ordinary' teaching, then the concept has no meaning.)
- What theories or benchmarks do teachers use when they make judgments and decisions?
- What do teachers look for when they are assessing? Are teachers' theories and benchmarks translated into criteria for sampling and interpreting student performance? If yes, how?

There is also a potential problem, first identified by Bernstein, that curriculum-based assessment activities might actually *disadvantage* social and cultural minority students further, because they rely on shared implicit knowledge of the tasks, criteria, and processes involved in the first instance – precisely the kinds of knowledge not necessarily readily available to such students. As Gipps (1999: 383) concludes: 'it is clear that social and cultural issues are just as significant in assessment at the classroom level as they are in external assessment at the system or societal level. What we look for in our students and whether we assess performance of the individual or the group are culturally determined'. Following from this, if more equitable, inclusive assessment practices are to be developed consistently for

ethnic minority students, Gipps (1999: 387) argues that a key direction for the future lies in the development of teachers' classroom assessment skills:

This implies the continued development of new assessment strategies for use by teachers, involving group and interactive assessment and interview and portfolio approaches. It will involve extending teachers' skills in observation and questioning while making them aware of social and cultural influences on the assessment process. We need to bring out into the open the nature of the power relationship in teaching and assessment and point out the possibility of reconstructing this relationship. Perhaps most important, we need to encourage teachers to bring pupils into the process of assessment, in order to recognize their social and cultural background, and into self-assessment, in order to develop their evaluative and metacognitive skills. All of these acts are ... both possible and necessary if assessment is to be more equitable and fulfil its promise to aid and support high quality learning.

In the New Zealand context, the principal challenge for encouraging more equitable, inclusive, teaching and assessment practices, would also clearly appear to be the upskilling of teacher knowledge and practice in these areas. Given that the student population is increasingly diverse and the teaching population increasingly homogeneous (as well as aging), this challenge remains a major one for New Zealand education and one that needs urgently to be addressed in preservice and continuing education. That the so-called 'literacy tail', and the over-representation of Māori, Pasifika and other ethnic minority students within it (see Section 2, below), has remained relatively unchanged over the last decades, confirms as much.

2. The language of assessment

This second section explores the international literature on the relationship between language(s) and assessment. A key concern of this literature is how monolingual-testing contexts can disadvantage minority, bilingual and/or ESOL students in relation to their monolingual peers. The principal reason identified in the literature for this differential pattern of achievement is that 'the language of the test' often confounds assessments of the academic ability of linguistic minority students.

This finding has particular implications for the New Zealand assessment context, given the consistent and significant 'home-language gap' (Wilkinson, 1998) that has been identified here over the last few decades in relation to international, comparative literacy surveys and its close relation to New Zealand's so-called 'literacy tail'. Briefly, the home-language gap is the gap between the literacy achievements of students whose home (or first) language corresponds with that of the school – and thus, by extension, national educational assessment practices – and those students for whom it does not. New Zealand has the *largest* home-language gap of any OECD country at both 9 and 14 years. In other words, New Zealand has the *poorest* performance across the OECD in the successful acquisition of English language literacy for those students who do not speak it as a first language, including many Pasifika, ESOL, immigrant and refugee students (for further discussion, see May, 2002a, b; Franken, May, & McComish, 2005). Addressing, and remediating, this ongoing gap in educational performance for such groups should thus be a priority in any evaluation of current assessment practices.

Specifically, the following key issues are identified as salient in the international literature.

Modular view of language

Most tests of language proficiency are underpinned by a modular view of language. According to this view, language is seen as comprising discrete elements, each related to a particular domain of language structure and skills. This perspective assumes that these domains are relatively autonomous and can be analyzed independent of the influence of cultural or extrinsic factors (cf. Section 1). Tests that reflect this modular view of language include, among others: the Boehm Test of Basic Concepts; the Peabody Picture Vocabulary Test; the Carrow Elicited Language Inventory.

However, as Gopaul-McNicol and Armour-Thomas (2002: 62) argue, these tests, intended for monolingual English speakers, are inappropriate for use with ESOL students, or what they term English language learners (ELLs), because they lack construct validity – such measures ignore or undervalue manifestations of language developed through different experiences within different cultural contexts.

As a consequence, these standardized language measures:

with their emphasis on the assessment of decontextualized language skills, are likely to favor children whose language socialization and acquisition were influenced by experiences within mainstream culture. In this sense, while reliable and valid for children socialized according to the norms and values of the mainstream English-speaking culture ..., these tests are invalid for children whose primary language socialization and acquisition occurred in other cultures. (ibid: 62-63)

Misdiagnosis of language difficulties

This leads, in turn, to the potential wider misdiagnosis of minority and/or ELL students as the result of monolingual assessment practices. As Gopaul-McNicol and Armour-Thomas (2002: 64-65) summarize it, there are two key factors that may contribute to such misdiagnosis:

- *Similarities in problematic behaviours*: ELLs demonstrate similar language behaviour at times to language impaired monolingual English speakers (e.g. grammatical errors, mistakes in phonology, comprehension difficulties)
- *Complexity of the process of second language acquisition*: Five features are identified here that may indicate inadequate language development in a monolingual speaker but are normal in second language acquisition
 - *Interlanguage* – an internal language system that combines the students' first language rules, English rules, and/or rules adapted from either or both languages
 - *Code switching* – the regular shifting between languages
 - *Rule fossilization* – maintenance of a less than perfect interlanguage system, reflecting the learner's stage of second language acquisition
 - *Linguistic interference* – misuse of vocabulary, errors in grammar, pronunciation difficulties
 - *Language loss* – loss of first language skills in the second language learning process because of lack of bilingual support.

And yet, these normal second language-learning processes are still regularly pathologized in monolingual assessment settings. As LaCelle-Peterson (2000: 28; my emphasis) also observes of this, 'despite clear recognition in the professional measurements standards that an individual's linguistic and cultural heritage *and current level of proficiency in the language of the measurement instrument* influences the meaning of obtained scores', misconceptualizations about the academic abilities of second language learners, and related educational misdiagnosis, persist.

As he proceeds to argue, if proficiency in the language of assessment varies across a test taking population, the assumption that test takers are essentially the same on all relevant features is violated. Amplifying this point, LaCelle-Peterson cites the *Standards for Educational and Psychological*, in its section on the testing of linguistic minorities: 'For a non-native English speaker and for a speaker of some dialects of English, every test given in English becomes, in part, a language or literacy test'

(American Psychological Association 1985: 73). This simple concession raises for him a critical, practical question: to what extent does a given student's score represent the target construct, and to what degree does it reflect proficiency in the language of assessment? And even that question, he suggests, oversimplifies matters when one remembers ELLs are not a single population sharing a single trait: they are a varied population with varied and constantly changing levels of proficiency.

From this, LaCelle-Peterson concludes starkly that when ELLs are involved in assessments conducted in English, the results yielded are not comparable to those of native English speakers and are also of uncertain comparability to those of other ELLs whose level of proficiency may differ along an often wide second language-learning continuum:

In other words, language proficiency constitutes a confounding variable that differentially influences the scores of students; varying language proficiency levels of ELLs compromise the validity and reliability of tests that seek to measure anything other than English language proficiency itself. Test scores that reflect a combination, in unknown proportion, of a target construct such as academic achievement in a particular subject area and second language proficiency in the test language are difficult to interpret at the individual level and, when aggregated into larger sets of test data, introduce errors which preclude making accurate evaluation of the schools or programmes serving the students. (ibid: 32-33)

These differential educational effects are clearly evident in recent academic analysis of No Child Left Behind (NCLB) in the USA, which requires all students, irrespective of their first language background, to be assessed in English for the purposes of national tracking and monitoring. Menken (2008) has documented the clearly negative educational effects that NCLB testing policies are having on English Language Learners (ELLs) in this regard in New York City. She found that the requirement that ELLs pass New York state exams in English in order to graduate from high school posed enormous challenges for both ELLs and their teachers. As a result, each year, scores attained by ELLs ranged from 20 to 50 percentage points below native English speakers.

Academic Language Proficiency

Another key contributing factor to the difficulties facing ELLs in relation to English language assessment practices relates to the central notion of academic language proficiency. First articulated by Jim Cummins in the 1970s, this notion has since garnered widespread support as a key explanatory variable in the educational success, or otherwise, of bilingual and/or ELL students (see Thomas & Collier, 2002; May, 2002a,b; May, Hill & Tiakiwai, 2004 for further discussion).

Briefly, the core argument concerning this notion is as follows. While it usually takes a second language learner only one to two years to learn a second language to a conversational standard, it may take as long as *four to eight years* of consistent high quality instruction in the second language for students to acquire the academic language proficiency necessary to reach grade-level norms and perform

intellectually challenging tasks such as those required to engage successfully in literacy activity. This difference in conversational and academic second language learning has been termed the 'second language learning delay' and is the result of ELLs having to learn new classroom-based language – which is demonstratively more demanding than conversational language – at the same time as new curriculum content in that academic language register.

The additional difficulties ELL students face in relation to assessment become clear in light of the second language learning delay, although they continue to be largely ignored by monolingual assessment practices, including those in New Zealand. As Lueng and Lewcowicz (2008) summarize it, ELL students, irrespective of their English language proficiency and schooling backgrounds, are expected to participate in standardized subject assessment, which has been devised with (only) native English speakers in mind. For those students who are still learning to use English for academic purposes effectively, the English language in standardized assessments can thus often pose an additional linguistic challenge that distorts their ability to demonstrate their content knowledge as effectively as their monolingual peers. Whenever this happens it would invalidate test scores as indicators of content knowledge and achievement.

One concrete example of this, discussed by Solono-Flores and Nelson-Barber, is when students who are tested in a second language have to deal with specialized knowledge within that second language that is specific to the discipline. As they outline, '[s]tudents who may understand a scientific phenomenon or a mathematical principle may still not be able to demonstrate that knowledge because of the lack of appropriate academic vocabulary. Even if they possess that vocabulary, the way in which they use it may not look or sound as technical as it would look or sound for native speakers' (2001: 559).

More broadly, Lueng and Lewcowicz conclude: 'all of this raises serious fundamental questions about the validity of using a set of non-differentiated criteria for the assessment of second language students' English and curriculum achievements' (2008: 306). Likewise, where students are required to participate in the full range of curriculum study and assessment, irrespective of their first language background, it is necessary to ask the following key language model- and construct-related questions:

- How should assessment deal with the relationship between curriculum content and classroom language use?
- What is English language proficiency in curriculum contexts? (ibid: 312)

Similar to the conclusions reached in Section 1, Lueng and Lewcowicz argue that the particular challenges that inhere for second language learners in mainstream (monolingual English) teaching and assessment contexts require teachers to have far greater knowledge and expertise than currently in the formative assessment of both English and content learning:

Not only do teachers need to be knowledgeable about models and trajectories of second language development in a curriculum context, they will also need to be familiar with the relationships between subject content and the language expressions associated with the subject content, and how such relationships bear on second language development. (ibid: 311)

Arguments related to the implementation of 'effective teaching practices', in general terms, though often purported to be the answer to issues of differential student achievement, do *not* address this issue. Effective teaching practices are a necessary, *but not sufficient* condition for teaching second language learners well. One actually also needs to know how to teach language and content effectively in these contexts – that is, language related pedagogical content knowledge (PCK) (see May & Wright, 2007 for further discussion). And yet, New Zealand teachers currently have very little opportunity in either their preservice training or continuing professional development to address *directly* key issues of second language development and the implications of these for teaching, learning and assessment. This might well be a key contributing factor in the ongoing home-language gap evident in New Zealand school achievement indices.

Translations

A common practice for mitigating the effects of a potential language mismatch in assessment practices is the use of translation(s). The most common approach to translating assessments is the translation-back-translation procedure. A text is translated from a source into a target language; a second interpreter (or group of interpreters) independently translates the text back into the source language. The accuracy of the translation is evaluated by comparing the original and back-translated versions (van de Vijver & Tanzer, 2004). While this procedure has been widely used, it can still lead to errors or imbalances. A good example of this, given by Hambleton (1994: 235), is the test item 'Where is a bird with webbed feet most likely to live?' The Swedish translation of 'bird with webbed feet' became 'bird with swimming feet', which provides a much stronger clue to the solution than the original item.

This example highlights the difficulties of equivalence between two languages. Ensuring equivalence between languages is not easy to establish because of within-language variations in terms of dialectal and idiomatic expressions and the difficulty of finding literal equivalents. Moreover the rules for ensuring that words have the same level of difficulty, same frequency of use, and the same meaning are difficult to maintain in translation (Gopaul-McNicol & Armour-Thomas, 2002).

The approach to the translation task will also depend on whether a new assessment instrument is to be developed or whether an existing instrument is to be translated. The former is known as simultaneous development and the latter as the successive development of different language versions. As van de Vijver and Tanzer (2004) argue, from a methodological perspective, the first option is often easier to carry out because typical problems of successive development, such as the use of local idioms

which are difficult to translate, can often be easily avoided. Still, the most common practice is to use successive development.

As van de Vijver and Tanzer (2004: 122-123) proceed to outline, three options are available to researchers in the successive development method. The first is *application*, which amounts to the literal translation of an instrument into a target language. In this option, it is assumed that the underlying construct is appropriate in each cultural group and that a simple, straightforward translation will suffice to get an instrument that adequately measures the same construct in the target group. The literal translation is by far the most common option in test translations.

The second option is *adaptation*. For some instruments, it is unrealistic to assume that a simple translation will yield an instrument that will adequately cover the same construct in the target group. An adaptation amounts to the literal translation of a part of the items and/or changes in other items and/or the creation of new items. Adaptations are based on the notion that the use of the application option would yield biased instruments. For example, a core of common items may show construct bias because they poorly sample the domain of possible items in at least one culture and, hence, the construct is insufficiently represented.

Finally, in some cases, the instrument has to be adapted to such a degree that practically a new instrument is assembled – hence, this third option is called *assembly*. In particular, when construct bias caused by differential appropriateness of the item content for the majority of the items threatens a direct comparison, assembly may be the best option. Another indication for using the assembly option would be the incomplete overlap of the construct definition across cultures (e.g. aspects of the construct that are salient for some cultures but are not covered in the instrument).

As van de Vijver and Tanzer conclude, it is clear that the three translation options differ in the amount of items that can be retained in the translation process. Going from the first to the third option, an increasing number of items will be changed in the translation process.

In light of the obvious challenges that inhere in the translation of assessment instruments, particularly with respect to obtaining or ensuring linguistic equivalence, it is clear that particular caution needs to be exercised with respect to such practices. To my knowledge, there is no research available in the New Zealand context with respect to translation of assessments for Pasifika, Asian and/or migrant/refugee students. However, there is some discussion of the challenges of translation with respect to Māori -medium education, which has clear relevance to other ethnic groups.

For example, as Rau (2008) outlines, one of the early solutions for addressing a shortage of assessment procedures in Māori was to translate standardized tests directly from English to Māori. Often carried out by novice test developers (mainly teachers) unaware that psychometric properties do not readily transfer or translate from one language to another, these tests quickly circulated to other Māori medium classrooms in other schools where further changes would often be made, seriously compromising the reliability and validity of the tests.

Other responses to demands for student achievement information in Māori -medium education include reconstructions of existing English language tests, which allow for more flexibilities and concessions than is possible with direct translations. Rau (2005a,b) reports the positive implications for Māori medium education of one such reconstruction, a procedure that is highly valued both nationally and internationally. This has generated robust literacy achievement data for students in Māori medium to boost the limited pool of information that currently exists. The simultaneous or delayed parallel development in English and Māori of new tests, are other alternatives that most typically characterize assessment development for Māori medium, although as Rau (2008: 326-327) argues, all of these

versions create the potential for inappropriate comparisons between the performances of students across the different language contexts to be made. With all of these scenarios, English medium defines for Māori medium what competencies will be tested and how they will be measured. It also means that at best, only partial representations of indigenous language structural and content domains are possible.

Consequently, Rau concludes, the option most preferred by educators in Māori medium is the development of original (new) tests and assessment frameworks sensitive to issues of second language acquisition, commensurate with a Māori worldview and consistent with the broader issues of Māori language and cultural revitalization. A few such literacy tests and literacy assessment frameworks have been commissioned by the Ministry of Education (for example see Rau, Whiu, Thomson, Glynn, and Milroy, 2001) but have not (yet) been promoted nationally by the Ministry of Education, for reasons that remain unclear.

One other key arena where the use of translation of assessment instruments into Māori has proved problematic is the National Education Monitoring Project (NEMP), which cyclically samples student performance in both English and Māori-medium contexts. In the early days, much energy and resources were expended in providing 'close' Māori translations of the original English language tasks. However, as we have seen, treatment, no matter how careful, cannot eliminate the inherent bias that places at a disadvantage those being tested using translated versions, particularly when the test is presented in the test taker's second language (as is the case for the majority of students in Māori-medium education).

As summarized by May, Hill and Tiakiwai (2004), these difficulties were acknowledged by NEMP in their 2001 report (National Educational Assessment Research Unit, 2001), where they highlighted the following key disparities between the English and Māori language assessment practices developed to that point:

1. The development and selection of some tasks may have advantaged the English-medium students, as mainstream teachers and researchers developed the majority of tasks.
2. The earlier assessments were translations of English texts with which the Māori -medium students may not have been familiar, and which may have also included unfamiliar dialectal vocabulary.
3. The activities in the Māori texts were often more complex than the English versions.

4. The students in the 1999 sample did not necessarily have stable Māori proficiency, as their te reo Māori abilities were not screened beforehand. Some may also have had only one or two years experience in Māori-medium contexts, thus also potentially disadvantaging them with respect to the assessment of grade appropriate material in te reo Māori (see the earlier discussions on the second language learning delay). In order to redress this, the second sample in 2000 included only those students with *at least 5 years* in Māori -medium education.
5. There are significant educational issues regarding the comparability of the English-medium and Māori -medium groups, given that Māori-medium education lacks resources and qualified teachers, something that Māori students in general English-medium schools would not experience.

As a result of these issues coming to prominence at that time, NEMP has since included tasks for learners in English medium that assess Māori language knowledge and more recently there has been significant movement toward the independent development of tasks in Māori (Rau, 2008).

3. New Zealand research on assessment for Pasifika, Asian, immigrant and refugee students

The first two sections have highlighted a range of key issues identified in the international literature in relation to the influence of culture and language on assessment practices. Following from this, one might expect such issues to be addressed in the New Zealand research literature on assessment. However, apart from nascent discussions in relation to Māori-medium assessment (see Section 2), there is currently an almost complete absence of research that addresses *directly* the implications of (monolingual English) assessment practices on other cultural and linguistic minority groups, particularly Pasifika, Asian, immigrant and refugee students. This is a worrying lacuna, particularly given that many from these groups are disproportionately represented in the lower echelons of educational achievement in New Zealand.

Educational achievement patterns

That said, there *is* increasing research discussion that focuses directly on the relatively poor educational achievements of ethnic minority students in New Zealand, particularly in light of international comparisons of literacy and numeracy achievement (e.g. SEA, PISA, PIRLS, IEA, TIMMS), and the significant home-language gap (see Section 2) identified by these comparisons.

This research highlights differences in achievement among and within different ethnic groups in New Zealand. As Biddulph, Biddulph and Biddulph (2003: 50) summarize the comparative educational achievement of these groups (see also, Alton-Lee, 2003; Hill & Hawk, 2000, 2003; Thomas & Kearney, 2008):

The picture of achievement [from TIMSS, NEMP, SEA, IEA and PISA] is remarkably consistent. That pattern is as follows: the achievement of the three groups Pakeha/European children, Asian children, and 'Other' non-European children is at a similar level and, for the most part, is significantly higher than that of Māori and Pasifika children. There are exceptions in the NEMP assessments in those cases where tasks developed to 'measure' particular knowledge and skills used Māori contexts (for example, a marae context in social studies, a kai moana context in science). On these tasks, Māori and Pasifika children performed significantly better than other children. Māori children's achievement is similar to that of Pasifika children in most cases. Where there is a difference between these two ethnic groupings, it is the Māori children who demonstrate higher achievement, at least through primary and middle schooling. The 2001 data show this pattern changes at senior secondary school, with a significantly greater proportion of Māori than Pasifika children leaving schooling early (63% of Māori are retained to age 16 compared to 84% of Pasifika children). This change is also evident in the number of children obtaining Year 12 qualifications.... only 41% of Māori leavers in 2001 had Sixth Form Certificate (or equivalent) or higher qualifications, compared with 55% of Pasifika leavers. This retention pattern is of particularly concern for Māori children in schooling.

Towards the end of their best evidence synthesis, Biddulph et al. conclude that these differences in achievement 'may be partly accounted for by assessment procedures that tend not to reflect Māori/Pasifika contexts for learning' (ibid: 178). This conclusion is reinforced by the NEMP findings discussed by Biddulph et al in the above quote, which highlighted how Māori and Pasifika students' performance increased when culturally relevant contexts were included in assessment practices.

Educational interventions

The differences in educational achievement, and, in particular, the worrying patterns of lower achievement for Māori and Pasifika students, has seen an emerging research focus on school-based interventions to 'improve' the educational performance of these groups. This nascent research has focused, in particular, on building research-informed communities of practice, aimed at improving teaching and learning practices within schools, and thus improving the educational achievement of minority groups. Raising teacher expectations, addressing deficit constructions of students, and encouraging critical discussions among teachers are key features of this work (see Millward, 2006 for a useful summary). Examples of this emergent research include, amongst others, *Picking up the Pace* (Phillips, McNaughton & MacDonald, 2001) and *Shifting the Focus: Achievement information for professional learning* (Timperley, 2003; see also Timperley & Wiseman, 2003).

While this is important research, with a particular focus on the educational achievement of cultural and linguistic minority students, it seldom addresses directly the issues of culture and language in relation to assessment, as highlighted in the first two sections of this report. The New Zealand research to date also fails largely to acknowledge how the language backgrounds of students impact on the teaching and learning process, and vice versa. Thus, much of the focus, in both research and intervention, has been on 'improving' English literacy for such students, disconnected from their home language backgrounds.

Culturally (and linguistically) relevant assessment

The studies discussed above, along with the impact of Te Kotahitanga for example (see Meyer, Weir, McClure, Walkey & McKenzie, 2007), have contributed significantly to raising awareness of the importance of culturally relevant pedagogy and practice in New Zealand schools. However, the ongoing invisibility of language background in discussions of educational achievement for cultural and linguistic minority students remains a cause for concern. In Alton-Lee's (2003) otherwise excellent best evidence synthesis on quality teaching for diverse students, there is, surprisingly, very little discussion of the role of language. One notable exception is when Alton-Lee describes a study of the writing process (O'Rourke & Philips, 1989, in Alton-Lee, 2003: 34), where the authors identify practices that were reported by teachers as being particularly effective for Māori and Pasifika students. These included using ethnically diverse role models, scaffolding of oral expression more

directly into written, facilitating peer support (teina/tuakana or tuakana/teina), and giving prestige to the oral as well as written 'publishing' of work.

Franken, May, and McComish (2005) highlight effective bilingual education practices for Pasifika students at individual schools such as Finlayson Park Primary School, Richmond Road Primary School and Clover Park Middle School in Auckland (see also McCaffery & Tuafuti, 2003; Tuafuti & McCaffery, 2005), which draw directly on the first language backgrounds of students in the teaching and learning process. While these examples accord with best practices for bilingual students in the international literature, they have not generated any momentum at the national level as yet, despite the fact that many Pasifika students have knowledge of a Pasifika language.

NCEA

The move to standards-based assessment in New Zealand, of which NCEA is the most prominent example, might suggest an improvement for cultural and linguistic minority students in the medium- to longer-term (given the advantages of these types of assessment, as discussed in Section 1). However, overall achievement patterns have not changed substantially for these students, at least as yet (McDonald & Thornley, 2005). This may be the result of the ongoing lack of teacher knowledge and expertise of second language teaching and learning principles (see Section 2) – a key factor in the effectiveness (or otherwise) of teacher's classroom-based assessment practices of these students in the NCEA process.

Participation patterns for ethnic minority students, though, may be changing as a result of the impact of NCEA. Fitzpatrick and Locke (2008) highlight, for example, how NCEA has had a positive impact on the ongoing academic participation of Māori and Pasifika students in low decile secondary schools, giving these students increased access (and, relatedly, more prestige) to previously low-status subjects such as Physical Education.

This relates to a point made, more broadly, by Thrupp (2006:9), when he suggests that assessment practices need to be re-thought in light of the wider imperative for 'curricular justice':

The main thing within schools is to disrupt forms of curriculum (and pedagogy and assessment) which privilege the cultures of some social and ethnic groups over others.... An example might be, say, a programme which provided an authentic history of a particular Pasifika culture, say Samoan culture, [which] was taught and assessed in a way which gave genuine weight to the language and culture of those students as it stands in New Zealand today, and where (crucially), the qualification gained was seen to be of equal standing to that gained in other kinds of courses.

Streaming

One other key area that continues to be a cause for concern with respect to assessment in relation to ethnic minority groups is the practice of using assessment to stream students.

Whitehead (2007) provides the following example of the potentially negative effects of using assessments for streaming purposes in the New Zealand context. He describes how one secondary school used a literacy 'entry examination', constructed by its English Department, for this purpose and administered to all Year 9 students. While literacy is clearly more than spelling and grammar, the 'entry examination' comprised: (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'. As Whitehead concludes:

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.

Whitehead's concern here relates to the wider trend identified in the international literature of the differential effects of streaming on cultural and linguistic minority students (see Section 1). New Zealand would appear to be little different in this respect, since many secondary schools continue to use assessments to stream students and there is a (perhaps not unrelated) over-representation of Māori and Pasifika students in lower bands or streams in New Zealand secondary schools.

4. Conclusions and Implications

This report has highlighted the key issues that impinge on the assessment of cultural and linguistic minority students in schools. Given the paucity of the New Zealand research in this area, much of report has necessarily focused on the wider trends and issues identified in the international literature, with respect to the cultural locatedness of assessment (Section 1) and the relationship between language and assessment (Section 2). Section 3 explored these issues in relation to the New Zealand education context. While there is currently very little research that directly explores the links between these issues in assessment and their implications for Pasifika, Asian, immigrant and refugee students, it is hoped that the report might provide a basis for useful extrapolation in the ongoing review of assessment practices in New Zealand.

Meanwhile, the following key issues need to be urgently addressed in any such review:

- The lack of New Zealand based research on the cultural and language- related dimensions of assessment practices
- The lack of teacher knowledge of, and professional development opportunities for, implementing effective assessment practices in culturally and linguistically diverse classrooms.
- The lack of second language learning and teaching knowledge as a means of informing those classroom-based assessment practices.

References

- Alton-Lee, A. (2003). *Quality teaching for diverse students in schooling: Best evidence synthesis*. Wellington: Ministry of Education.
- American Psychological Association (1985). *Standards for Educational and Psychological Testing*. Washington DC.: American Psychological Association.
- Biddulph, F., Biddulph, J., Biddulph, C. (2003). *The complexity of community and family influences on children's achievement in New Zealand: Best evidence synthesis*. Wellington: Ministry of Education.
- Black, P. (2000). Research and the development of educational assessment. *Oxford Review of Education*, 26(3/4), 409-419.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-73.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education *Assessment in Education*, 11(1), 7-26.
- Crooks, T. (2002). Educational assessment in New Zealand schools. *Assessment in Education*, 9(2), 237-253.
- Delandshere, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational achievement. *Assessment in Education: Principles, Policy and Practice*, 8(2), 113-133.
- Fitzpatrick, K., & Locke, T. (2008). Sir, do we get credits for this? NCEA, class and ethnicity in New Zealand. In S. Brown, J. O'Neill & A. St George (Eds.), *Facing the big questions in education: Purpose, power and learning*. Melbourne: Thomson.
- Fox, J. (2008). Alternative assessment. In E. Shohamy & N. Hornberger (eds), *Language Testing and Assessment. Encyclopedia of Language and Education (2nd edition)*, Volume 7 (pp. 97-109). New York: Springer.
- Franken, M., May, S., and McComish, J. (2005) Pasifika Languages Research Guidelines Project. Literature Review. Hamilton: Wilf Malcolm Institute of Educational Research, University of Waikato. February 2005.
- Fusarelli, L. D. (2004). The potential impact of the no child left behind act on equity and diversity in American education. *Education Policy*, 18(January and March), 71-94.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24, 355-392.
- Gopaul-McNicol, S., & Armour-Thomas, E. (2002). *Assessment and culture: Psychological tests with minority cultures*. San Diego, CA: Academic Press.
- Hambleton, R.K., 1994. Guidelines for adapting educational and psychological tests: a progress report. *European Journal of Psychological Assessment (Bulletin of the International Test Commission)* 10, 229-244.
- Hawk, K., Tumama Cowley, E., Hill, J., & Sutherland S. (2001). Relationships: The critical factor in teaching Maori and Pasifika students. Paper presented to the New Zealand Association for Educational Research (NZARE) Annual Conference, Christchurch, December 2001.

- Hill, J., & Hawk, K. (2000). *Making a difference in the classroom: Effective teaching practice in low decile, multi-cultural schools*. Wellington: Ministry of Education.
- Hill, J., & Hawk, K. (2003). *Achieving is cool: What we learned from AIMHi project to help schools more effectively meet the needs of their students*. Auckland: IPDER Massey University.
- LaCelle-Peterson, M. (2000). Choosing not to know. Assessment policies and practices obscure the education of language minority students. In A. Filer (Ed.), *Assessment: Social practice and social product* (pp. 27-42). London: RoutledgeFalmer.
- Lueng, C., & Lewcowicz, J. (2008). Assessing second/additional language of diverse populations. In E. Shohamy & N. Hornberger (Eds.), *Language Testing and Assessment. Encyclopedia of Language and Education (2nd edition)*, Volume 7 (pp. 301-317). New York: Springer.
- May, S. (2002a). Where to from here? Charting a way forward for language and education policy in Aotearoa/New Zealand. *The TESOLANZ Journal*, 10, 22-35.
- May, S. (2002b). The problem with literacy. *English in Aotearoa*, 48, 4-14.
- May, S. (2007). Sustaining effective literacy practices over time in secondary schools: school organisational and change issues. *Language and Education* 21(5), 387-405.
- May, S., Hill, R., & Tiakiwai, S. (2004). *Bilingual/immersion education: indicators of good practice. Final Report to the Ministry of Education*. Wellington: Ministry of Education.
- May, S., & Wright, N. (2007). Secondary literacy across the curriculum: challenges and possibilities. *Language and Education* 21(5), 370-376.
- McCaffery, J., & Tuafuti, P. (2003). Samoan children's bilingual language and literacy development. In R. Barnard & T. Glynn (Eds.), *Bilingual children's language and literacy development* (pp 80-107). Clevedon, England: Multilingual Matters.
- McDonald, T., & Thornley, C. (2005). *Literacy teaching and learning during the secondary years: Pathways for success*. Paper presented at the Literacy Without Boundaries: 14th European Conference on Reading, Zagreb, Croatia.
- Menken, K. (2008). High-stakes tests as de facto language policies. In E. Shohamy & N. Hornberger (Eds.), *Language Testing and Assessment. Encyclopedia of Language and Education (2nd edition)*, Volume 7 (pp. 401-413). New York: Springer.
- Meyer, L. H., Weir, K., McClure, J., Walkey, F., & McKenzie, L. (2007). *Final report: Longitudinal research on the relationship between the NCEA and student motivation and achievement*. Wellington: Jessie Hetherington Centre for Educational Research, Victoria University.
- Millward, P. (2006). *Organisational learning facilitated by the analysis of student achievement information*. Unpublished Education Doctorate, University of Auckland, Auckland.
- Moss, P., & Schultz, A. (2001). Educational standards, assessment, and the search for consensus. *American Educational Research Journal*, 38, 37-70.
- National Educational Assessment Research Unit. (2001). *Assessment Results for Maori Students 2000: Music, Aspects of technology, Reading and Speaking*. Dunedin: National Educational Assessment Research Unit.
- Okazaki, S., & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. *Psychological Assessment*, 7(3), 367-375.

- Phillips, G., McNaughton, S., & MacDonald, S. (2001). *Picking up the Pace - Effective literacy interventions for accelerated progress over the transition into decile 1 schools: Final report to the Ministry of Education on the Professional Development associated with the Early Childhood Primary Links via Literacy (ECPL) Project*. Auckland: The Child Literacy Foundation and Woolf Fisher Research Centre.
- Rau, C. (2005a) Literacy acquisition: Assessment and achievement of year two students in total immersion in Māori programmes. *International Journal of Bilingual Education and Bilingualism* 8, 5,404-432.
- Rau, C. (2005b). Te Hāpai Ako, Mātauranga Reo, Kura Strand of Pen Schooling Improvement Initiative, Report to the Ministry of Education, Ngāruawāhia.
- Rau, C. (2008). Assessment in indigenous language programmes. In E. Shohamy & N. Hornberger (Eds.), *Language Testing and Assessment. Encyclopedia of Language and Education (2nd edition)*, Volume 7 (pp. 319-330). New York: Springer.
- Rau, C., Whiu, I., Thomson, H., Glynn, T., and Milroy, W. (2001). *He Ara Angitu. A description of success in reading for five-year old Maori medium students*. Wellington: Ministry of Education.
- Shepard, L. A. (2000). The role of assessment in learning culture. *Educational Researcher*, 29(7), 4-14.
- Solono-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of research in science teaching*, 38(5), 553-573.
- Timperley, H. (2003). *Shifting the focus: Achievement information for professional learning*. Wellington: Ministry of Education.
- Timperley, H., & Wiseman, J. (2003). *The Sustainability of Professional Development in Literacy PART TWO School-Based factors Associated with High Student Achievement*. Wellington: Ministry of Education.
- Thomas, W. and Collier, V. (2002). *A National Study of School Effectiveness for Language Minority Students' Long-Term Academic Achievement*. Santa Cruz, Ca.: Center for Research on Education, Diversity and Excellence (CREDE).
- Thomas, S., & Kearney, J. (2008). Teachers working in culturally diverse classrooms: Implications for the development of professional standards and for teacher education. *Asia-Pacific Journal of Teacher Education*, 36(2), 105-120.
- Thrupp, M. (2006). *Improving the schooling chances of New Zealand's poorest children: Policy and community challenges*. Paper presented at the AGM of the Child Poverty Action Group, Auckland.
- Tuafuti, P., & McCaffery, J. (2005). Family and community empowerment through bilingual education. *The International Journal of Bilingual Education and Bilingualism*, 8(5), 480-503.
- Tuck, B. (1995). Issues of objectivity in assessment: A plea for moderation. In R. Peddie & B. Tuck (Eds.), *Setting the standards: The assessment of competence in national qualifications* (pp. 59-78). Palmerston North: Dunmore.
- van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue europeenne de psychologie appliquee*, 54, 119-135.
- Whitehead, D. (2007). Literacy assessment practices: Moving from standardised to ecologically valid assessments in secondary schools. *Language and Education* 21(5).

Wilkinson, I. (1998). Dealing with diversity: achievement gaps in reading literacy among New Zealand students. *Reading Research Quarterly*, 33(2), 144-167.