'One assessment doesn't serve all the purposes' or does it? New Zealand teachers describe assessment and feedback

S. Earl Irving · Lois R. Harris · Elizabeth R. Peterson

Abstract Within the Asia-Pacific community, the New Zealand Ministry of Education has been one of few educational authorities to adopt an Assessment for Learning (AfL) framework and actively promote formative uses of assessment. This paper reports the results of a qualitative study in which eleven New Zealand secondary teachers in two focus groups discussed their conceptions of assessment and feedback. These data were examined to see how teachers defined and understood assessment and feedback processes to identify how these conceptions related to AfL perspectives on assessment. Categorical analysis of these data found teachers identified three types of assessment (formative, classroom teacher–controlled summative and external summative) with three distinct purposes (improvement, reporting and compliance, irrelevance). Feedback was seen as being about learning, grades and marks, or behaviour and effort; these types served the same purposes as assessment with the addition of an encouragement purpose. This study showed that although these New Zealand teachers appeared committed to AfL, there was still disagreement amongst teachers as to what practices could be deemed formative and how to best implement these types of assessment. Additionally, even in this relatively low-stakes environment, they noted tension between improvement and accountability purposes for assessment.

Keywords Teacher conceptions of assessment · Teacher conceptions of feedback · New Zealand · Secondary school

S. E. Irving (✉) · L. R. Harris
Faculty of Education, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
e-mail: e.irving@auckland.ac.nz

L. R. Harris
e-mail: lois.harris@auckland.ac.nz

E. R. Peterson
Department of Psychology, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
e-mail: e.peterson@auckland.ac.nz

Introduction

In the last two decades, work by authors including Sadler (1989), Black and Wiliam (1998a, b) and the Assessment Reform Group (e.g. Assessment Reform Group 1999, 2002) has advocated that teachers should adopt an *Assessment for Learning* (AfL) perspective towards assessment. The Assessment Reform Group (1999) noted that:

> A clear distinction should be made between *assessment of learning* for the purposes of grading and reporting, which has its own well-established procedures, and *assessment for learning* which calls for different priorities, new procedures and a new commitment. (p. 2)

When defining AfL, the Assessment Reform Group (1999) explained that AfL differed from traditional understandings of assessment in that it:

- is embedded in a view of teaching and learning of which it is an essential part;
- involves sharing learning goals with pupils;
- aims to help pupils to know and to recognise the standards they are aiming for;
- involves pupils in self-assessment;
- provides feedback that leads to pupils recognising their next steps and how to take them;
- is underpinned by confidence that every student can improve;

- involves both teachers and pupils reviewing and reflecting on assessment data (p. 7).

Hence, AfL differs from traditional assessment practices in that it advocates that students should be actively involved in the assessment and evaluation of their own work through formative assessment (e.g. Black et al. 2003; Weeden et al. 2002). Formative assessment can be defined as: 'assessment carried out during the instructional process for the purpose of improving teaching or learning' (Shepard 2006, p. 627) and is often contrasted with summative assessment, which is 'assessment carried out at the end of an instructional unit or course of study for the purpose of giving grades or otherwise certifying student proficiency' (Shepard, p. 627).

New Zealand educational policy has adopted the discourse of AfL, with the Ministry of Education actively promoting formative assessment through policy, resources, and professional development (Brown and Harris 2009; Crooks 2002; Ministry of Education 2004). For example, in a recent position statement on assessment, the New Zealand Ministry of Education stated:

> For some years now, our approach to assessment has been moving beyond a narrow summative ('end point' testing) focus to a broader focus on assessment as a means of improving teaching and learning. This is sometimes referred to as *assessment for learning*…. In New Zealand, our approach is very different from that in other countries. We have a deliberate focus on the use of professional teacher judgment underpinned by *assessment for learning* principles rather than a narrow testing regime. Professional judgment about the quality of student work is guided by the use of examples, such as annotated exemplars, that illustrate in a concrete way what different levels of achievement look like (2010, p. 5).

This stance makes New Zealand one of the few countries in the Asia-Pacific region to openly adopt an AfL assessment philosophy, although a number of Western European countries have similar assessment philosophies. New Zealand schools are self-governing, and although teachers are bound by a National Curriculum and, since 2010, National Standards,[1] teachers and school administrators can choose how and when they combine informal assessment methods, teacher-created summative assessments and teacher-administered standardised assessment tools [e.g. Assessment Tools for Teaching and Learning (asTTle), Progressive Achievement Tests (PAT)] to evaluate student learning. It is only during the final 3 years of secondary

school (Years 11–13) when students encounter an external standard-based qualification system [National Certificate of Educational Achievement (NCEA)] made up of internally and externally assessed components.[2]

The relatively low-stakes assessment environment in New Zealand provides a distinct contrast to the majority of its Asia-Pacific neighbours, generally freeing teachers from the external assessment pressures often discussed in the United States and the United Kingdom (e.g. Amrein-Beardsley et al. 2010; Black and Wiliam 2005; Kruger et al. 2007; Yeh 2006). Koh and Luke (2009) point out that many Asian countries, such as Hong Kong, Taiwan, Japan, Singapore and South Korea, are beginning the process of delinking themselves from their traditionally strong examination culture and exploring the feasibility of formative practices, therefore heightening interest in how nearby countries like New Zealand are progressing with these reforms.

While the New Zealand policy context is considered conducive to an AfL agenda, some research suggests that AfL objectives are only being partially realised due to the narrow, and at times ineffective, implementation of formative practices (e.g. Harris 2009; Hume and Coll 2009). It is likely that AfL's objectives will be achieved only if teachers and other stakeholders understand and are fully committed to implementing AfL principles. Hence, it is important to identify teacher thinking, as this underpins their practices (Fishbein and Ajzen 1975).

This study explored the relationship between New Zealand secondary teachers' conceptions of assessment and feedback, with the purpose of gaining a better knowledge of how teachers view various types of assessment and their purposes and how they see feedback as related to or distinct from the assessment process. The paper begins by discussing existing literature on teacher conceptions of assessment and feedback, demonstrating how these relate to theoretical best practice conceptions. It then reports data from a qualitative study of New Zealand secondary teachers and concludes by examining the wider implications these data hold for the implementation of AfL in the Asia-Pacific region.

Teacher conceptions of assessment and feedback: the literature

Studies into teacher conceptions of assessment and feedback are important because they give insight into the complex ways teachers view, interpret and ultimately act in their role as assessors and providers of feedback. While there is a growing body of work examining teachers'

---

[1] National Standards were introduced in 2010 for the first 8 years of compulsory schooling and therefore do not apply to secondary schooling. The study in this paper occurred before National Standards were introduced.

[2] Students can obtain one of four grades for NCEA–Not Achieved, Achieved, Merit or Excellence. These terms are also used by the teachers in this study when referring to their classroom assessments.

conceptions of assessment, few studies have looked at how they understand feedback. This is surprising given Hattie's (2009) synthesis of meta-analyses on the effects of schooling found that feedback was one of the most powerful moderators that enhanced achievement, consistent with the AfL discourse.

Within this paper, assessment and feedback are considered to have a cyclical relationship; an assessment normally occurs *before* feedback can be given about performance. While assessment can occur without feedback being given to the learner, feedback is based on some form of assessment. So, while assessment and feedback are highly related, within this paper, they will be discussed as separate constructs. This review first examines and defines the term 'conception' as it features heavily in the literature about teacher conceptions of assessment and feedback. It then discusses teacher conceptions of assessment and feedback as reported in the literature, comparing and contrasting these with theoretical models.

Conceptions

Much of the literature around teacher understandings of assessment and feedback examines their conceptions of these key processes, making it important to first examine what we mean by the terms concept and conception. Thagard (1992) notes that the traditional view in cognitive psychology is that a concept is a group of objects or behaviours that are defined in such a way that they can be used and widely recognised. This type of definition has also been referred to as 'concepts as prototypes'. This traditional definition suggests that a concept is relatively stable and unchanging and ignores the different contextual, personal and affective components that often become attached to an individual's understanding of concepts. For this reason, Entwistle and Peterson (2004) make the distinction between a 'concept' which is something that is largely mutually understood and a 'conception' which is an individual's personal and changeable response to a concept.

Researchers in education interested in concepts and conceptions have typically taken one of the two approaches (Tynjala 1997). The cognitive approach has tended to focus on people's mental models of concepts (often in the physical sciences and mathematics) and has investigated how students can change or replace their naive conceptions for conceptions that are more technically and scientifically correct (Vosniadou and Brewer 1992). A key component of conceptual change is metacognitive or metaconceptual awareness which is argued to be a precursor for any major conceptual change or restructuring (Vosniadou 2007). The other approach is a phenomenographic approach that focuses more on the variety of ways that people understand

and describe concepts and the impact this has on their behaviour and in particular on the way that people teach and learn (Entwistle and Peterson 2004; Lowyck et al. 2004; Marton and Booth 1997; Peterson et al. 2010).

In this paper, the focus is on how teachers describe and understand the concepts of assessment and feedback, and hence, the approach aligns more with the phenomenographic research in this area. While teachers are likely to be aware of a textbook definition of assessment and feedback, their response to these constructs are personal in that they develop through their own experiences of different educational contexts and situations. As such, their conceptions of assessment and feedback are likely to be varied, to have an emotional component and to potentially hold inherent contradictions. Contradictions in people's conceptions are thought to arise because their conceptions are frequently partial or incomplete (often because they are unstated) or because their conceptions of a phenomenon can change as they move from one context to another (Marton and Pong 2005; Halldén 1999). Hence, when examining people's conceptions, others may identify seeming inconsistencies that the holder of the conception does not view as inherently problematic.

Finally, in trying to comprehend someone's conceptions, it is useful to understand what they think something is as well as what it is not. This is because concept formation requires the ability to discriminate between responses that are inappropriate or irrelevant to a particular situation (Kendler 1961).

Overall, an awareness of one's conceptions is argued to be a precursor for any major conceptual change or restructuring (Vosniadou 2007). This exploration of a sample of secondary teachers' conceptions of assessment and feedback and their relationship to each other may also serve as a base for other researchers interested in promoting conceptual change.

Conceptions of assessment

Within an Asia-Pacific context, Brown and colleagues (e.g. Brown 2004, 2008; Brown et al. 2009; Brown and Lake 2006; Harris and Brown 2009) have conducted the most robust quantitative and qualitative research into teacher conceptions of the purposes of assessment. Confirmatory factor analyses with samples in New Zealand, Queensland (Australia) and Hong Kong have demonstrated an acceptable fit of data to three purposes of assessment (Brown 2007, 2008; Shohamy 2001; Torrance and Pryor 1998):

- it improves teaching and student learning (improvement);
- it makes students accountable for their learning (student accountability);

- it makes schools and teachers accountable for student learning (school accountability).

Additionally, an anti-purpose and fourth conception has been identified:

- assessment should be rejected because it is invalid, irrelevant and negatively affects teachers, students, curriculum and teaching (irrelevance).

The improvement conception is most aligned with AfL principles (Black et al. 2003) and posits that assessment's primary role should be to improve student learning and to a lesser extent improve teachers' teaching.

For students, improvement may occur directly through the provision of feedback and effective peer and self-assessment practices or indirectly through teacher modification of instruction to better suit diagnosed student needs. Both teacher-based intuitive judgment and formal assessment tools can be used credibly for this purpose. For teachers, improvement occurs by using assessment results as a basis from which to modify their instruction and improve their teaching.

The final three conceptions are more aligned with evaluative assessment purposes. The student accountability conception posits that assessment holds students responsible for their own learning through the assigning of grades and scores and through external examinations and qualifications; this information can then be reported to community stakeholders like parents, other schools and employers. The school accountability conception suggests that assessment can publicly demonstrate teacher and school effectiveness through formal often standardised assessment (Firestone et al. 2004). The irrelevance conception reflects feelings that formal evaluation has no legitimate place within schooling when it diverts time away from teaching and learning, is unfair or negative for students or is perceived to be invalid or unreliable.

Across a number of countries and sectors, teacher questionnaire responses identified that they generally agreed with both improvement and student accountability conceptions and disagreed with school accountability and irrelevance conceptions (Brown and Ngan 2010). Harris and Brown (2009) found seven conceptions within their qualitative data on New Zealand teachers' conceptions that fit broadly into three purposes: improvement, accountability and ignore/reject. In Harris and Brown's study, teachers discussed considerable tension between improvement and accountability purposes, noting that accountability purposes sometimes forced teachers to use assessment in ways that they felt would not lead to student improvement.

Within an AfL framework, assessment's fundamental role is to lead to student improvement (Assessment Reform Group 1999). Advocates of AfL have suggested a range of reforms to substantially reduce or eliminate traditional forms of testing and grading (Assessment Reform Group 2002), citing the negative effects it can have on student self-esteem and motivation to learn. Instead, alternative assessment forms like comment-only marking, peer-assessment and self-assessment are suggested (Black et al. 2003). However, it remains unclear how these alternative kinds of assessments could be used to completely replace more formal evaluations for the purposes of certifying and improving learning.

Conceptions of feedback

While there is a growing body of literature examining student conceptions of feedback (e.g. Carnell 2000; Lipnevich et al. 2008; Peterson and Irving 2008; Poulos and Mahony 2008), very little research has been conducted on teachers' conceptions of feedback. One of the few studies is Brown et al. (2010). This study trialled a 7-item *Teachers' Conceptions of Feedback* inventory with a national sample of New Zealand primary and secondary school teachers ($N = 518$). Within this sample, teachers most strongly agreed with factors relating to learning-orientated feedback and rejected grading-oriented factors. A positive finding was that teachers within this sample did not endorse factors suggesting feedback should be used to promote student well-being, meaning that teachers generally agreed that feedback needed to be about academic performance, not effort or behaviour.

While research on teachers' conceptions of feedback is scarce, theoretical and empirically based models suggest that there are a range of types of feedback with differing purposes or outcomes (e.g. Askew and Lodge 2000; Butler and Winne 1995; Hargreaves 2005; Hattie and Timperley 2007; Shute 2008; Tunstall and Gipps 1996). For example, Tunstall and Gipps' (1996) work in the United Kingdom found that teachers gave four types of feedback relating to socialisation and management (performance orientation), rewarding and punishing (performance orientation), specifying attainment and improvement (mastery orientation) and constructing achievement and the way forward (learning orientation). While teachers utilised all types of feedback, they most frequently used their preferred types. Tunstall and Gipps argued that their first two categories led to a performance orientation towards learning, the third to a mastery orientation and the fourth to mastery and constructivist perspectives (a learning orientation).

Hattie and Timperley's (2007) review of feedback literature identified four types of feedback and factors mediating their effectiveness. These were feedback task (e.g. whether work was correct or incorrect), feedback process (e.g. comments about the processes or strategies underpinning the task), feedback self-regulation (e.g. reminding students of

Table 1 Participant demographic information

Participant	Sex	Subject taught	Years teaching	Ethnicity	School socio-economic level
Annette	Female	Mathematics	5–10	European	Middle
Bert	Male	Science	Greater than 10	European	Middle
Caroline	Female	English	Greater than 10	European	High
Deborah	Female	Mathematics	Greater than 10	European	Middle
Emma-Jayne	Female	English	Less than 5	Pacific Islander	Low
Fatima	Female	Mathematics	Greater than 10	Middle Eastern	Low
Harvey	Male	English	Greater than 10	European	Middle
Justin	Male	Mathematics	5–10	European	High
Kelly	Female	English as a second language	Greater than 10	European	Low
Mark	Male	English	Greater than 10	European	Middle
Rachel	Female	English	Less than 5	European	Middle

strategies they can use to improve their own work) and feedback self (e.g. non-specific praise and comments about effort). The review found that of these, feedback self is least effective as it does not provide information about to how to improve. Feedback on process and self-regulation is considered most powerful (Butler and Winne 1995), but task feedback is reported as most common.

Research methods

The review in the previous section has identified that while there is considerable research on teacher conceptions of assessment, little has been done to examine their conceptions of feedback. This study aimed at increasing knowledge about how teachers understand assessment and feedback, looking at how they see these processes as discrete and related.

Research questions

The following research questions guided this study.

1. How do New Zealand secondary teachers understand assessment and feedback?
2. How did these conceptions relate to each other?

Sample

Participants for this study were secondary school teachers recruited from schools taking part in the Conceptions of Assessment and Feedback (CAF) project, a larger 2-year study that focussed on teachers' conceptions of assessment and feedback in Year 9 and Year 10 classrooms and ways in which their conceptions impacted on classroom practices (For information about the project, see Peterson et al.

2006). Teachers in secondary schools were of interest because in New Zealand they conduct a mix of internal, school-based assessment (for Years 9 and 10) followed by a national qualification system, NCEA (for Years 11–13). CAF project participants were recruited at the school level. Once school consent was obtained from the principal, teaching staff were invited to participate. Two to three teachers in each school volunteered to work on an individual action research project with one of their classes. Thus, the participants were a convenience sample of urban secondary school teachers.

Eleven teachers (7 women, 4 men; 4 mathematics, 6 English and 1 science) from four large, co-educational Auckland secondary schools participated in this study (see Table 1 for participant demographics). All but four participants had been teaching for more than 10 years. Nine were of European descent, with the remaining two reporting as Pacific Islander and Middle Eastern, respectively. Nationally, approximately 72% of teachers are women; ethnically, 75% of New Zealand teachers claim European descent, 3% are of Pacific Island descent and 2% are classified as Other, with the remainder identifying as either Maori or Asian (Ministry of Education 2005). The students in the schools came from diverse socio-economic backgrounds; one school was in a working class area of Auckland, one was from a socio-economically advantaged suburb, and the other two were in middle-class suburbs. Hence, while this was primarily a convenience sample, it was proportionally similar to the New Zealand teaching population and represented schools catering for a diverse range of students.

Data collection and analysis

Focus groups were chosen for data collection because they can be used to explore how participants view their lived

experiences (Kitzinger 1995; Krueger 1994). In addition, they can explore and examine the interactions amongst participants not found in interviews or questionnaires (Morgan 1988). Focus groups have previously been useful when exploring student conceptions of assessment and feedback (e.g. Lipnevich et al. 2008; Peterson and Irving 2008; Poulos and Mahony 2008).

The eleven participants were divided into two groups and each participated in a 90-min focus group conducted by a member of the research team. These focus groups took place at the beginning of the 2-year project and were the first opportunity for the participants to formally express and discuss their views on assessment and feedback. Where possible, teachers from the same school were separated in an effort to maximise their freedom of response. The proceedings of the focus groups were audio-recorded and transcribed verbatim.

The focus groups explored three key aspects of assessment and feedback: definition, purpose and personal response. As a common problem in focus groups is engaging reluctant participants (Fontana and Frey 2000), prior to any discussion within the group, teachers were asked to write definitions of assessment and feedback, their purposes and a personal response to each on coloured Post-It ® notes/stickies. They placed these on the wall under the headings of Definitions, Purposes and Personal Response (see Peterson and Barron 2007, for a description of this process). The stickies were then used to initiate discussion within the group and were later collected for textual analysis and triangulation with the transcripts.

Coffey and Atkinson's (1996) categorical analysis technique was used to analyse data. A mixture of a priori and emergent coding was utilised. As the discussion had been structured around three categories (i.e. definitions, purposes and personal response), these became the overarching a priori categorisations for analysis. After all authors had read the transcripts multiple times, each was involved in the process of assigning data to the three a priori categories (i.e. definitions, purposes and personal response); statements were assigned to multiple categories when appropriate.

Once participant statements had been identified that related to the three categories, a process of emergent analytic coding was used where subcategories were allowed to 'grow out of' or 'emerge' from the data (Lankshear and Knobel 2004) and these were named using participants' actual words. Each author examined the data set independently and proposed codes that they saw as appropriately describing the participant perspectives. These codes were then compared and contrasted iteratively with codes generated by the other authors to create a set of categories that aptly described the data. All authors reviewed the final categorisation of the data and reached consensus on classifications.

Results

Teachers within this study saw clear relationships between assessment and feedback, with similar purposes articulated for each. Teachers described formative, teacher-controlled summative and external summative assessments as being related to three purposes: student improvement, reporting and compliance, and irrelevance or no purpose. Feedback was divided into feedback on learning, feedback on behaviour, and grades and marks; these could be for encouragement, improvement, or reporting and compliance purposes, or for no purpose (in the case of providing comments with grades, teachers believed that students focused solely on the grades and ignored the comments, thus rendering the comments irrelevant). In this section, categorisations will be outlined and then illustrated using data from the study.

Conceptions of assessment

Teacher descriptions of assessment types were underpinned by three distinct purposes. While teachers wanted assessment to be for student improvement purposes, they saw reporting and compliance as another required and often legitimate use of assessment. Additionally, teachers articulated that some assessments served no purpose whatsoever as data were either ignored or rejected as invalid by teachers or students; that is, these assessments became irrelevant to the teachers even though other stakeholders likely saw legitimate purposes for them. Teachers described strategic use of assessment, identifying that different types often served differing purposes. As Fatima noted, 'If it's for a report, you have to use national standards[3]; that's how it makes it reliable. If you're using [it] for the students, it's different; one assessment does not serve all the purposes'.

Participants classified actual assessments into formative, classroom teacher–controlled summative and external summative assessment categories. Teachers related formative assessment most strongly to their preferred assessment purpose, student improvement. Formative assessment was described as non-graded teacher, student and peer interactions which provided diagnostic information that could direct the teacher's instruction and/or provide feedback to the student about how to improve their work.

The final two types of assessment were most strongly related to reporting and compliance purposes, with the distinction between the two primarily related to locus of

---

[3] This teacher is referring to the need for teachers to use reliable nation-wide assessment tools to benchmark their reports, and not to the National Standards that were introduced in 2010 for the primary school years.

control. Classroom teacher–controlled summative assessment was described as graded or marked work that could provide a reportable snapshot of student achievement; these assessment events were created or selected by the students' actual teacher to best align with the classroom learning intentions. External summative assessments were described as tests or test-like practices selected by school leaders (such as a department head or school principal) or external agencies. Teachers described external summative as distinct from their own summative uses of assessment as they did not have input into the way these assessments were selected and conducted, leading to potential mismatches between the assessment, student abilities and what had been taught. The relationships between these types of assessment and their purposes are described in Fig. 1. Within this model, solid lines indicate clear agreement and endorsement amongst the teachers and dotted lines represent pathways where there was more moderate endorsement and agreement amongst the participants.

Improvement purposes

The improvement purpose was the teachers' most preferred reason for assessing students, and formative assessment was described as the best way to provide useful assessment information to students. For example, one teacher wrote on a stickie, 'I LOVE formative assessment. It empowers me and students.' However, teachers differed in what they considered to be formative practices:

Emma-Jayne: … your formative assessment is testing that you do throughout this unit. So I'll be doing little tests throughout the unit,

testing what they know. For example, I do cloze activities and things like that, just testing their comprehension instead of not really knowing what they know, so formative is just little tests you do on the way.

Rachel: A lot of formative for me is assessing prior knowledge, seeing what they know already so I'm not going to bore them stupid because of the assumption that they don't know how to paragraph and they do, something like that.

Caroline: Formative for me… it's like your formative years. It's when you're allowed to make mistakes … I learn through mistakes. That's basically my only way of learning, when I've made a boob [colloquial for 'mistake'] and I've been told 'it's okay this time, but next time we won't be so easy on you.'

Kelly: And I think also part of the process of formative is to give feedback. Formative assessment is not only to find out where they've got to [go] or what they understand, but as you say, to actually help them to know what they need to learn next as well.

The formative types discussed here include diagnostic assessments to determine prior knowledge, tests and activities to assess progress throughout a unit, the giving and receiving of feedback and opportunities for students to learn from mistakes. While teachers like Emma-Jayne saw

Fig. 1 Teacher understandings of assessment
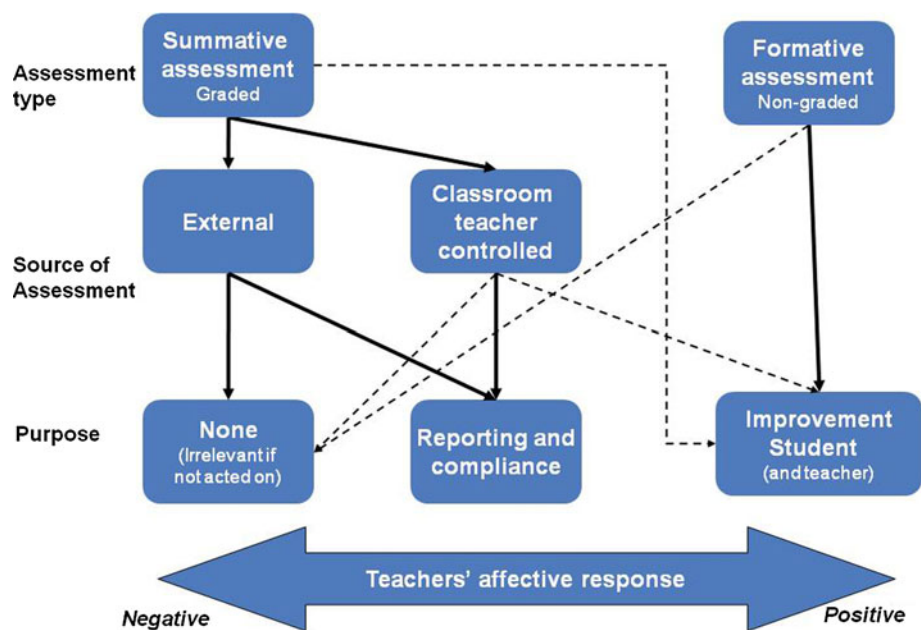
testing as potentially being formative, these tests were described as 'little' and diagnostic; most teachers in this study did not describe tests as formative.

Most of the comments about the improvement focus of formative assessment were centred on using formative assessment to help students understand how to improve. However, some teachers also acknowledged that it could be used to improve and 'inform our teaching' (stickie). As one teacher explained:

Emma-Jayne: Even as teachers you do self-assessments on yourself … we're always assessing ourselves. We don't seem to think it but we do. If we find that there's some gaps in a group of kids then you think, 'Okay, how could I have done that? Maybe I need to re-look at the way I taught that and teach it in a different way.'

Hence, while the primary focus was providing information for students to use to improve their learning, teachers also reported using these data to improve their teaching.

Reporting and compliance purposes

Teachers were less positive about the uses of summative assessment as they described these as relating primarily to reporting and compliance purposes. One teacher noted on a stickie, 'I don't enjoy summative assessments. It feels like the end'. Summative assessment was defined on stickies as: 'A 'convenient' measurement for reports/exams'; 'A snapshot of progress in learning against some preset goal or criteria'; and 'Assessment gives a level and a position on a rubric for a student'. These descriptions are not inherently negative but are more strongly related to reporting than student improvement. The presence or absence of a mark or grade was described as the main difference between formative and summative assessment, with teachers aware that students were keenly interested in knowing whether an assessment would 'count towards our grades' (Deborah). 'To have grades' was a major reason for conducting these summative assessments, as they were 'about reporting, reporting to parents, reporting to the community if you like. This is a snapshot people; this is where we're at with our kids' (Harvey).

Some teachers were quite conscious of the tension created between summative assessment and the improvement purposes they were promoting through formative uses. For example, Annette explained

I'm in conflict professionally with myself…. I have to give them summative tests every term and report back to parents so although I'm trying to create one

culture in the classroom [i.e. a formative culture], when push comes to shove, I'm reporting back and give them marks that they are failures.

However, some teachers did see summative assessment and improvement purposes being potentially aligned. For example, within the conversation below, while Caroline clearly saw grades as the end of a feedback cycle, Harvey believed that grades could facilitate conversations with students that led to improvement:

Caroline: I'm talking about maybe internal assessment. Final assessment or graded. Even my junior classes, this is the grade that needs to go on your report; there's no room for manoeuvring after that, so they're not actually interested. I mean I might do that same assessment at the end of the year, but they're not actually interested in that at the moment because this is the grade that's recorded.

Harvey: It depends on what conception they've got of their achievement. Like I've had conversations with kids saying, 'Why didn't I get a Merit?[4] I thought it was worth Merit.' And I'll go through the criteria with them and talk about [it] and then they'll understand what their [Achieved] comes from.

Caroline: You'll only have two students doing that though won't you?.

Harvey: You won't get a lot, but you'll get a few.

While Harvey acknowledged that few students used grades in formative ways, he did describe this as possible unlike most teachers in this sample. The majority questioned the validity and reliability of grades awarded to students. As Rachel pointed out 'it's still bandied about by teachers of what descriptors mean', making it very difficult for teachers to mark consistently across schools and classes and for students to understand criteria in ways that would allow them to use grades as constructive feedback about their work.

Irrelevance of some assessments

While teachers generally described their own use of summative assessments (i.e. classroom teacher–controlled) more negatively than their formative practices, they still preferred these to external summative assessments like NCEA and department-wide testing. Often teachers failed to see legitimate purposes for these external assessments, describing them as irrelevant to student learning. For example, one teacher wrote on a stickie, 'Much happier designing my own assessments. Marking using current

---

[4] Merit is a specific grade within the NCEA framework, sitting between Achieved and Excellence.

NCEA methods has affected me extremely negatively (and the students too)'. Another wrote: 'Assessment should be fair, transparent. Irrelevant if too external'. Even though this study encouraged teachers to discuss their assessment and feedback practices within low-stakes Year 9 and 10 contexts, the NCEA programme clearly dominated teacher thinking about assessment and had obvious washback effects. Teachers attempted to prepare students for this assessment regime by introducing them to the language and structures of NCEA style assessments prior to Year 11, evidenced by teacher discourses within the data around NCEA grade terminology and assignment structure.

One of the reasons teachers wished to ignore or reject external summative assessment was because it was seen as damaging to students' self-esteems. As Annette noted:

I enjoy watching them [students] progress and … even when I put the tests on the table, I know that they're going to fail and they will be defined as failures… it almost feels that it's what happens in the classroom versus what happens in the school and the department. It's internal/external and it's almost like me and the students versus the rest… it's watching their faces as you encourage and their self-esteem grows and grows and grows and it comes near the end of term and we're all getting tense because we know that this test is going to happen and we're all going to fail.

Annette's remarks demonstrated the way that some teachers within the sample constructed a dichotomy dividing personal classroom assessment practices from those externally required. This passage also highlighted how teachers in the sample saw themselves as unsuccessful when students did poorly on external assessment, using the pronoun 'we' to include themselves in the failure.

External summative assessment was also rejected or deemed irrelevant by some because it was seen as eroding confidence in teacher judgements.

Annette: … quite often within the department or within the structure of a school, you're asked to give a result by a pass or a fail by a test, assuming that my teacher assessment or the student–teacher interaction of assessment is not valid enough because it hasn't been by a formal test … I think we've got… to hold onto our professional judgments.
Justin: So for that you do need the reliability of staff.
Annette: That's what people externally will always say, 'How reliable? How can you show that it's reliable?' Because if everybody gets the same test, then we can say it is reliable.

Within this discussion, Justin's statement indicates that some teachers do accept external monitoring as necessary because of inconsistencies between teachers. However, multiple teachers did raise issues around the reliability of test data, highlighting the many factors that might negatively affect student performance on the day (e.g. student mood, physical conditions and comprehension of the questions asked), preventing a 'true' result from being obtained through assessments which were not ongoing.

All types of assessments were considered as potentially being irrelevant if they were not used to drive improvement. For example,

Kelly: I think that's a fair point that if you do assess either you've got to change your teaching according to what you find out or you've got to ignore the assessment completely and teach what the programme says you've got to teach.
Caroline: And I could do that, but I'm not feeling very happy about that.
Kelly: No, that's right and I back you up. I think that's a dilemma you have when you're assessing is that whatever information you get you then either have to act on it and that can change things or you have to ignore it and then there wasn't any point in doing the assessment.

Hence, while external assessment was discussed most negatively by the teachers, they acknowledged that results from their own teacher-created assessments could also be ignored, making those also potentially irrelevant.

In summary, teachers reported three main purposes for assessment. All teachers articulated that formative assessment was useful for the purpose of student improvement. While all agreed that classroom teacher–controlled and external summative assessments were for accountability purposes, within the group, there were some teachers who saw these as potentially inaccurate, making these data irrelevant. Few teachers reported using summative results for formative purposes or seeing their students using data from grades to improve their learning. While teachers said that expectations and criteria should be explicit and transparent, they indicated that this is seldom the case and highlighted the negative emotions that summative assessment results could generate for students.

Conceptions of feedback

Feedback was regarded as information about student learning that was provided after a formal or informal assessment event. Teacher descriptions of feedback types were underpinned by four distinct purposes for the feedback. While teachers valued improvement and encouragement purposes for feedback, they acknowledged that some feedback was given because of stakeholder expectations or to hint to students what kind of grade they could expect on

their final assignment. Teachers explained that the provision of comments alongside grades served no purpose because students focused only on the grade. Participants described three types of feedback: information about learning, grades or scores and comments on behaviour and effort. Verbal or written comments about learning or behaviour/effort were seen as having the most direct effect on student learning.

Figure 2 shows the relationships teachers described between feedback types, their purposes and learning. The purpose of the feedback was driven largely by the purpose of the preceding type of assessment (note the purpose of assessment and purpose of feedback labels in Figs. 1 and 2 are very similar, with the addition of an encouragement purpose for feedback). Finally, depending on the purpose of the feedback, different feedback content was given to the students. Solid lines represent a strong consensus amongst the participants, while dotted lines represent pathways where opinions diverged.

All four feedback purposes—improvement, reporting, irrelevance and encouragement—and the tensions among them (in the teachers' minds)—are captured in the following conversation.

Kelly: Most of my feedback would be oral feedback… I always put it [reading running records results] onto a graph and I always show them the graph to show the improvement that they're making. So that gives us a chance to talk about sort of where they're at … It's always discussion, so I suppose that's giving a mark and a comment.

Caroline: I'm very clear about not putting a grade on formative assessment, but I'm really uncomfortable about not putting a comment on summative assessment even though I know most of the time it's pointless and I mainly do it with the kids who don't achieve,… trying to pick the kids' self-esteem up and I don't know whether it works or not… I would just find it hard to [just] put a Not Achieved, even though I know that's probably the most efficient use of my time and say even as I'm handing it back to say, 'Hey, you nearly did', but it's a politeness thing.

Kelly: I know it's time consuming, but I think even somebody who's got an Excellence might well benefit from knowing what was excellent about it. 'Which of the things that I did are the things that actually achieved Excellence?' because then they can do it again.

Harvey: For formative assessment, what I do is I write a long comment. And I bury the grade it would get if it was being summatively assessed within the comment so they actually have to wade through the comment to get to the grade…. because to me, you need to give them an idea of where they're heading towards with that piece of work, particularly at senior level.

Here, teachers talked about feedback used for improvement, citing how students can use comments to enhance their performance, consistent with AfL perspectives. Some comments were designed to encourage students to keep trying; other teachers articulated that marks and grades were useful references to expected standards. Teachers did not agree that it was appropriate to include

Fig. 2 Teacher understandings of feedback
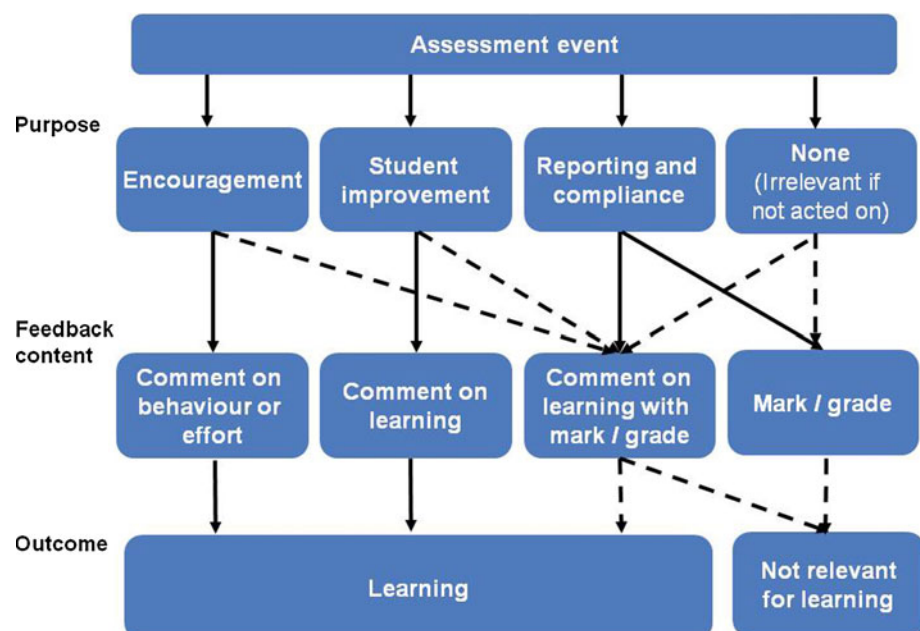
grades along with comments; as Harvey mentioned, by embedding a 'ballpark' grade in his comment, he indicated to the student (and other stakeholders) what to expect, showing this feedback was 'to keep parents and students informed of progress and needs' (Stickie).

In keeping with formative practices, teachers talked primarily about giving feedback on learning. They described a range of different practices that would all be examples of learning feedback:

Justin: It's all about showing the student where they've got to, where they need to go to next … Again I think being able to sit down with a student and talk to them would be ideal, but more often it is just a note on an assessment or some work that they've done just pointing out what they've got wrong and how they can get that bit correct.

Deborah: If we take assessment as also watching what students are doing from a textbook or whatever, then … you're giving them feedback as you walk around …. I have to admit that when I'm marking tests that it's more like 'well, I'm sure they will notice that they haven't done this right' [I make a mental note] rather than physically writing on their paper that 'this is an area that you still need to improve on.' But you do talk to students when you give back the test and it [a common mistake] appears with my teaching.

This mixture of oral and written feedback is seen as letting students know what they need to work on in order to improve, reinforcing 'that kids do progress because we give them specific feedback' (Bert).

Teachers also talked about feedback designed to improve students' self-esteem and to acknowledge and encourage positive classroom behaviours. As Mark explained, 'We give feedback about behaviour: 'Oh, that's great the way you put your hand up or whatever'. Feedback about behaviour and effort was designed 'to encourage students' (Stickie). Low-achieving students were described as needing this type of feedback 'because their self-esteem is somewhat low', causing teachers to focus on 'how well they've done, how good it looked …' (Annette). Teachers noted that being able to report on what students can do rather than being forced to refer to grade-level appropriate criteria would be helpful because:

What your kids seem to need is rather than say 'you've worked hard, but you've failed', to say 'you've worked hard and this is where you are …. (you have) achieved at Level One, you know, you're

wonderful' rather than say 'I'm going to test you at Level Four and you've failed again.' (Kelly)

Teachers said they tried to find positive things to discuss in student work rather than always relating the student's work to grade-level appropriate criteria.

There was disagreement within the sample about whether or not grades were feedback. Harvey suggested:

The grades are shorthand for feedback aren't they? Well, I think they are. Students, parents and schools etc., so this is the kind of 'letting people know' side of things and this [teacher comments on work] is diagnostic and informative, well diagnostic and figuring out the 'what to do next' side of things I suppose.

Caroline, however, rejected the notion that grades could be embedded within formative feedback:

…you would never put a grade on formative assessment because that's what they would look at and they would not be as interested in your comments. I find that if I just put comments on it… then they're happy to look at them and internalise them. If you put a grade on, that's what kids are conditioned to focus on… there's no point in giving feedback on summative assessment; they do not want to know why they boobed if that was their last chance. The only time you can be, I feel, really productive or constructive with your comments is when its formative.

Most teachers agreed that on graded summative work, written comments were unlikely to lead to improved learning because students were likely to ignore them and focus solely on the mark or grade.

In summary, the teachers described four purposes and three types of feedback. Feedback about learning was seen as leading to student improvement unless it was accompanied by grades or marks. In that case, most teachers described it as being irrelevant because students only took notice of the mark or grade. Grades and marks were seen as a performance indicator for students, useful for reporting purposes. Teachers also noted that especially with low-performing students, encouraging feedback about behaviour and effort was important in promoting student engagement and resilience.

Discussion

This study examined a small sample of secondary New Zealand teachers' views on assessment and feedback. Teachers indicated that the purposes of both assessment

and feedback can be seen in one of the three main ways. Assessment and feedback:

- improve student learning and potentially teaching;
- report student performance to stakeholders in compliance with school and ministry regulations;
- can be detrimental or irrelevant to student learning.

With feedback, encouragement was regarded as an additional purpose.

The teachers distinguished between formative and summative assessment on the basis of whether a student's assessed work was graded or not. Non-graded work was generally regarded as formative, and written or spoken feedback was described as fostering student improvement and positive affective consequences for pupils. Graded assessments were viewed as summative assessments and split into two subcategories depending on the locus of control. If control lay with the classroom teacher (i.e. teacher-controlled summative assessments), while reporting and compliance was often the primary purpose, student improvement was also seen as possible; these assessments could also be considered irrelevant if they were not acted upon. Where control lay with a body outside of the classroom (e.g. school administration or external agencies), then there were strong associations with irrelevance or reporting and compliance purposes.

Teachers within this sample had a very positive mindset towards formative assessment, but at the same time recognised that there were external demands for summative assessments that referenced student learning to benchmarks or standards. They also expressed the view that external assessments were often detrimental to the students' self-esteems and sense of personal growth and achievement. This created a tension because, on the one hand, they acknowledged the need for reporting, but had difficulty conceptualising how this could be done without describing student achievement in ways which might be damaging to student engagement, motivation and self-esteem.

Results of this study were mainly consistent with Brown's (2008) large-scale research on teacher conceptions of assessment. Teachers within this study did cite the improvement, school accountability and irrelevance conceptions but did not indicate that assessment should make students accountable for their learning. These teachers reported a high level of ownership of their students' successes and failures, frequently using inclusive personal pronouns like 'we', 'us' and 'our' when describing student results. The tension between improvement and accountability purposes discussed in this study has also been noted by other authors (e.g. Harris and Brown 2009) and appears consistent with previous findings.

Likewise, there were major similarities between this study's findings about teacher conceptions of feedback and the results from Brown et al.'s (2010) survey of New Zealand teachers' conceptions of feedback. Both groups highly endorsed the notion that feedback should lead to improved student learning and most rejected the idea that grades were a legitimate form of feedback. However, teachers in the current study generally stated that feedback about behaviour and effort was important and led indirectly to student improvement, whereas Brown et al.'s sample showed a much lower level of endorsement of this kind of feedback. The viewpoint found in the current study that feedback about behaviour and effort leads to student improvement is at odds with Hattie and Timperley's (2007) review which identified that feedback self was '…rarely converted into more engagement, commitment to the learning goals, enhanced self-efficacy, or understanding of the task' (p. 96). At present, it is unclear why teachers in this study described this kind of feedback as effective despite a large body of research suggesting otherwise.

The presence of tension between improvement and accountability purposes, even within a relatively low-stakes assessment system, gives pause for advocates of AfL and indicates that more work must be done to find ways to resolve the conflict between accountability and improvement purposes of assessment. Some teachers in this study did seem to view formative and summative assessment as dichotomous and mutually exclusive, something authors like Hargreaves (2005) note is unproductive.

These data indicate that more work may be necessary to help teachers understand how 'summative' practices can be used in 'formative' ways to improve student learning instead of viewing them as purely methods for reporting and achieving compliance. To avoid assessments becoming irrelevant, teachers need to be encouraged to learn from them and use them as a basis for modifying and improving their instruction.

Conclusion

New Zealand has been at the forefront within the Asia-Pacific community in relation to its widespread adoption of the AfL agenda and its lack of external high-stakes testing during the first 10 years of schooling. While the results from this study are based on a small sample and therefore care must be taken in generalising the findings, they nevertheless contain some interesting and worthwhile findings. On the whole, data from this study provide some positive news for those promoting AfL. Unanimously, these teachers described formative assessment in positive ways and related it primarily to improved student learning. However, their descriptions of formative practices did differ. There was a division over whether tests and grades or marks could be used formatively. Hence, while teachers

appear to agree with AfL in principle, these data would suggest large variability in the ways formative assessments were viewed and conducted. These differing conceptions may help to explain the diverse and at times ineffective implementations of AfL that Harris (2009) and Hume and Coll (2009) have discussed in studies of New Zealand practice, reinforcing Argyris and Schon's (1974) distinction between theories of action and theories in action.

Additionally, even in this relatively low-stakes system, these New Zealand teachers note tension between improvement and accountability purposes for assessment. As Bryant and Carless (2010) have found, assessment practices are deeply bound in the local culture, and these teachers were constrained by the reality of their here-and-now. Even though assessment in Years 9 and 10 was low stakes, the Years 11–13 high-stakes assessments clearly influenced how assessment was practised in these year levels (e.g. grading using Not Achieved, Achieved, Merit and Excellence). How to maintain student and teacher focus on formative assessment practices despite the washback effect from such exams remains unclear. This tension between improvement and accountability purposes is likely to occur in other countries trying to make the shift from a rigid examination-based system to greater use of formative assessment practices. These findings suggest that countries that are making or contemplating making those changes need to be cognisant of the work required to come up with practical ways of overcoming these tensions that are culturally appropriate for their contexts.

References

Amrein-Beardsley, A., Berliner, D. C., & Rideau, S. (2010). Cheating in the first, second, and third degree: Educators' responses to high-stakes testing. *Education Policy Analysis Archives, 18*(14).

Argyris, C., & Schon, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass Publishers.

Askew, S., & Lodge, C. (2000). Gifts, ping-pong and loops—linking feedback and learning. In S. Askew (Ed.), *Feedback for learning* (pp. 1–17). London: Routledge.

Assessment Reform Group. (1999). Assessment for learning: Beyond the black box. Retrieved December 3, 2010, from http://www.assessment-reform-group.org/publications.html.

Assessment Reform Group. (2002). Testing, motivation, and learning. Retrieved March 27, 2009, from http://www.assessment-reform-group.org.uk.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, UK: Open University Press.

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

Black, P., & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. London: Department of Education and Professional Studies, King's College London.

Black, P., & Wiliam, D. (2005). Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *Curriculum Journal, 16*(2), 249–261.

Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education, 11*(3), 301–318.

Brown, G. T. L. (2007 December). *Teachers' conceptions of assessment: Comparing measurement models for primary & secondary teachers in New Zealand*. Paper presented at the New Zealand Association for Research in Education (NZARE) annual conference, Christchurch, NZ.

Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science Publishers.

Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of Multidisciplinary Evaluation, 6*(12), 68–91.

Brown, G. T. L., Harris, L. R., & Harnett, J. A. (2010). *Teachers' conceptions of feedback: Results from a national sample of New Zealand teachers*. Paper presented at the International Testing Commission 7th Biannual Conference, Hong Kong.

Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy and Practice, 16*(3), 347–363.

Brown, G. T. L., & Lake, R. (2006, November). *Queensland teachers' conceptions of teaching, learning, curriculum and assessment: Comparisons with New Zealand teachers*. Paper presented at the Annual Conference of the Australian Association for Research in Education (AARE), Adelaide, Australia.

Brown, G. T. L., & Ngan, M. Y. (2010). *Contemporary educational assessment: Practices, principles and policies*. Singapore: Pearson Education South East Asia.

Bryant, D. A., & Carless, D. R. (2010). Peer assessment in a test-dominated setting: Empowering, boring or facilitating examination preparation? *Educational Research for Policy and Practice, 9*(1), 3–15.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281.

Carnell, E. (2000). Dialogue, discussion and feedback - views of secondary school students on how others help their learning. In S. Askew (Ed.), *Feedback for learning* (pp. 46–61). London: Routledge.

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies*. London: Sage.

Crooks, T. J. (2002). Educational assessment in New Zealand schools. *Assessment in Education: Principles Policy & Practice, 9*(2), 237–253.

Entwistle, N. J., & Peterson, E. R. (2004). Conceptions of learning and knowledge in higher education: Relationships with study behaviour and influences of learning environments. *International Journal of Educational Research, 41*, 407–428.

Firestone, W. A., Schorr, R. Y., & Monfils, L. F. (Eds.). (2004). *The ambiguity of teaching to the test: Standards, assessment, and educational reform*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fontana, A., & Frey, J. H. (2000). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 645–672). London: Sage.

Halldén, O. (1999). Conceptual change and contextualisation. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 53–65). Oxford: Pergamon.

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education, 35*(2), 213–224.

Harris, L. R. (2009). *Lessons from New Zealand: Practical problems with implementing formative assessment practices within the primary and secondary classroom*. Paper presented at the Invited Curriculum and Instruction Research Seminar at Hong Kong Institute of Education.

Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy, & Practice, 16*(3), 365–381.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hume, A., & Coll, R. K. (2009). Assessment of learning, for learning, and as learning: New Zealand case studies. *Assessment in Education: Principles, Policy & Practice, 16*(3), 269–290.

Kendler, T. S. (1961). Concept formation. *Annual Review of Psychology, 12*, 447–472.

Kitzinger, J. (1995). Qualitative research: introducing focus groups. *British Medical Journal, 311*, 299–302.

Koh, K., & Luke, A. (2009). Authentic and conventional assessment in Singapore schools: An empirical study of teacher assignments and student work. *Assessment in Education: Principles, Policy & Practice, 16*(3), 291–318.

Krueger, R. A. (1994). *Focus groups: A practical guide for applied research*. Newbury Park: Sage.

Kruger, L. J., Wandle, C., & Struzziero, J. (2007). Coping with the stress of high stakes testing. *Journal of Applied School Psychology, 23*(2), 109–128.

Lankshear, C., & Knobel, M. (2004). *A handbook for teacher research*. Berkshire: Open University Press.

Lipnevich, A. A., Smith, J. K., & Barnhart, S. (2008, March). *Students' perspectives on the effectiveness of the differential feedback messages*. Paper presented at the American Educational Research Association Annual Meeting, New York City.

Lowyck, J., Elen, J., & Clarebout, G. (2004). Instructional conceptions: Analysis from an instructional design perspective. *International Journal of Educational Research, 41*, 429–444.

Marton, F., & Booth, S. (1997). *Learning and awareness*. Mahwah, NJ: Lawrence Erlbaum Associates.

Marton, F., & Pong, W. Y. (2005). On the unit of description in phenomenography. *Higher Education Research and Development, 24*(4), 335–348.

Ministry of Education. (2004). Using assessment tools. *Curriculum Update* (54). Retrieved March 3, 2009, from http://www.tki.org.nz/r/governance/curric_updates/curr_update54_e.php.

Ministry of Education. (2005). Report on the Findings of the 2004 Teacher Census. Retrieved December 15, 2010 from http://www.educationcounts.govt.nz/publications/schooling/teacher_census.

Ministry of Education. (2010). Ministry of Education position paper: Assessment [Schooling Sector]: Ko te Wharangi Takotoranga Arunga, a te Tauhuhu o te Matauranga, te matekitenga. Retrieved December 4, 2010, from http://www.minedu.govt.nz/theMinistry/PublicationsAndResources/AssessmentPositionPaper.aspx.

Morgan, D. L. (1988). *Focus groups as qualitative research* (Vol. 16). Newbury Park, CA: Sage Publications.

Peterson, E. R., & Barron, K. A. (2007). How to get focus groups talking: New ideas that will stick. *International Journal of Qualitative Methods, 6*(3), 140–144.

Peterson, E. R., Brown, G. T. L., & Irving, S. E. (2010). Secondary school students' conceptions of learning and their relationship to achievement. *Learning and Individual Differences, 20*, 167–176.

Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction, 18*, 238–250.

Peterson, E. R., Irving, S. E., Brown, G. T. L., Haigh, M., & Dixon, H. (2006). *Conceptions of Assessment and Feedback Project: TLRI Final Report*. Auckland, NZ: The University of Auckland.

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: the students' perspective. *Assessment & Evaluation in Higher Education, 33*(2), 143–154.

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

Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (pp. 623–646). Westport, CT: Praeger.

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow, UK: Pearson Education.

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Buckingham, UK: Open University Press.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22*(4), 389–404.

Tynjala, P. (1997). Developing education students' conceptions of the learning process in different learning environments. *Learning and Instruction, 7*, 277–292.

Vosniadou, S. (2007). Conceptual change and education. *Human Development, 50*, 47–54.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth - A study of conceptual change in childhood. *Cognitive Psychology, 24*, 535–585.

Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment: What's in it for schools?* London: Routledge Falmer.

Yeh, S. S. (2006). High-stakes testing: Can rapid assessment reduce the pressure? *Teachers College Record, 108*(4), 621–661.