Secondary teachers’ conceptions of the purpose of assessment and feedback

Lois Harris, University of Auckland
S Earl Irving, Cognition Consulting Ltd, Auckland
Elizabeth Peterson, University of Auckland

Paper presented to the Australian Association for Research in Education (AARE) Annual Conference December 2008, Brisbane, Australia

Correspondence concerning this paper should be addressed to Dr. Lois Harris, School of Teaching, Learning, and Development, Faculty of Education, The University of Auckland, Private Bag 92019, Auckland NZ 1142 or by email to lois.harris@auckland.ac.nz

Acknowledgement. This study was completed in part through the financial support of the New Zealand Council for Educational Research’s (NZCER) Teaching Learning and Research Initiative Grant to the ‘Conceptions of Assessment and Feedback Project’. The cooperation and participation of the teachers and schools in this project is greatly appreciated.
Secondary teachers’ conceptions of the purpose of assessment and feedback

Lois Harris, University of Auckland
S Earl Irving, Cognition Consulting Ltd, Auckland
Elizabeth Peterson, University of Auckland

Introduction
Teachers’ thinking influences their classroom practices (Clark & Peterson, 1986; Pajares, 1992; Thompson, 1992). In turn, teacher actions significantly impact pupil learning (Muijs, 2006; Muijs & Reynolds, 2005). While a large body of research has already examined teacher conceptions of teaching and learning (Bolhuis & Voeten, 2004; Boulton-Lewis, Smith, McCrindle, Burnett, & Campbell, 2001; Kember, 1997; Kember & Gow, 1994; Kember & Kwan, 2000; Samuelowicz & Bain, 2001; Trigwell, Prosser, & Waterhouse, 1999), comparatively little has focused on teacher conceptions of assessment and feedback, which are important and distinct parts of the teaching and learning cycle. This paper presents results from an empirical study examining how teachers understand these key educational concepts.

Conceptions of assessment
Assessment is considered to be one of the key processes in the teaching and learning cycle as it allows stakeholders to evaluate learning and potentially use this information to improve learning and instruction. Many academics divide assessment into two types: formative and summative. While there are different interpretations of these two types of assessment, within this paper, formative assessment is considered to be “…assessment carried out during the instructional process for the purpose of improving teaching and learning,” while summative assessment is “assessments carried out at the end of an instruction unit or course of study for the purpose of giving grades or otherwise certifying student proficiency” (Shepard, 2006, p. 627). Hence, the key difference between formative and summative assessment is in its timing (during instruction or after instruction) and its uses (to improve teaching and learning or for reporting purposes). Researchers are now noting that presenting formative and summative assessment practices as dichotomous is unproductive as many assessments serve multiple purposes (Hargreaves, 2005). While Croft, Strafford, and Mapa’s (2000) research demonstrated that the majority of primary teachers surveyed reported frequently or always altering their teaching as a result of information from standardized tests and diagnostic tools, an example of so-called ‘summative’ assessments being used for ‘formative’ purposes, more research is needed to determine the extent to which teachers’ conceptions allow them to see assessments as having multiple purposes.

There has been robust quantitative measurement conducted through questionnaires on teacher conceptions of assessment (i.e., G. Brown, 2002, 2004, 2006, 2007; G. Brown, Kennedy, Fok, Chan, & Yu, in press; G. Brown & Lake, 2006; Hargreaves, 2005; Philippou & Christou, 1997), but one weakness in this body of research is the lack of qualitative data on teacher conceptions; it is this weakness that the study reported in this
paper seeks to address. The most robust empirical research into teachers’ conceptions of assessment has been conducted by G. Brown and colleagues (G. Brown, 2002, 2004, 2006, 2007; G. Brown, Kennedy, Fok, Chan, & Yu, in press; G. Brown & Lake, 2006). G. Brown’s questionnaire on teachers’ conceptions of assessment has been used with New Zealand, Australian, and Hong Kong primary and secondary teachers and confirmatory factor analyses have demonstrated an acceptable fit of data obtained from these teachers to four purposes of assessment (G. Brown, 2004; Heaton, 1975; Shohamy, 2001; Torrance & Pryor, 1998; Warren & Nisbet, 1999; Webb, 1992):

- It improves teaching and learning.
- It makes students accountable for learning.
- It makes schools and teachers accountable for student learning.
- It should be rejected because it is invalid, irrelevant, and negatively affects teachers, students, curriculum, and teaching.

The first conception is commonly linked in the literature with formative assessment practices. The improvement conception, promoted by researchers like Black, Wiliam, and colleagues under the rubric ‘assessment for learning’ (Black & Wiliam, 1998; Black, Harrison, Lee, Marshall, & Wiliam, 2003), is underpinned by the premise that assessment’s primary role should be for student improvement. This may occur directly by giving students useful feedback and using 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.

The final three conceptions are most commonly associated with assessment practices more commonly referred to as summative. The second conception is based on the premise that assessment holds students personally accountable for their learning through the assigning of grades, scores, and external exams and qualifications. For this purpose, summative or external assessment is used to formally determine individual student abilities and achievements; this information about students can then be reported to community stakeholders like parents, other schools, and employers. The third conception is based on the notion that assessment should be used to publicly demonstrate teacher and school effectiveness (Firestone, Schorr, & Monfills, 1998). This conception also relies on the use of formal, usually standardized assessment, so results can be compared between individual schools and school districts. The final conception reflects feelings that formal evaluation has no legitimate place within teaching and learning. Assessment may be deemed irrelevant if it is seen as diverting time and attention away from teaching and learning, if it viewed as unfair or negative for students, or if it is viewed as invalid or unreliable.

Work by G. Brown and colleagues (G. Brown, 2002, 2004, 2006, 2007; G. Brown, Kennedy, Fok, Chan, & Yu, in press; G. Brown & Lake, 2006) has identified that across a number of countries and sectors, teachers consistently agreed with both improvement and student accountability conceptions and disagreed with school accountability and irrelevance conceptions. While this large-scale research is important, it remains unclear
why this agreement and disagreement is occurring. Further qualitative research is needed to better understand these response patterns.

**Conceptions of feedback**
While feedback is generally conceptualised as part of the assessment cycle, many academics argue it is worth examining in its own right. Hattie’s (2007) synthesis of meta-analyses on the effects of schooling found that feedback is the most powerful single moderator that enhances achievement, agreeing with work by Black, Wiliam, and colleagues (Black & Wiliam, 1998; Black, Harrison, Lee, Marshall, & Wiliam, 2003). Nicol and Macfarlane’s (2006) review of literature identified that good feedback practices:

1. help clarify what good performance is (goals, criteria, expected standards);
2. facilitate the development of self-assessment (reflection) in learning;
3. deliver high quality information to students about their learning;
4. encourage teacher and peer dialogue around learning;
5. encourage positive motivational beliefs and self-esteem;
6. provide opportunities to close the gap between current and desired performance;
7. provide information to teachers that can be used to help shape teaching. (pg 205)

Research has shown that teachers give a range of types of feedback; these are not seen as having the same purposes or outcomes. There are many feedback models cited in research literature (i.e. Askew & Lodge, 2000; Butler & Winne, 1995; Hargreaves, 2005; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006; Shute, 2008; Tunstall & Gipps, 1996). For example, Tunstall and Gipps (1996) created an often cited typology of feedback practices. They identified through empirical research that teachers of young children give four types of feedback relating to: socialization and management (performance orientation), rewarding and punishing (performance orientation), specifying attainment and improvement (mastery orientation), and constructing achievement and the way forward (learning orientation). In their study, all the teachers utilised all types of feedback, but had preferred types that they used most often. The argued that their first two categories lead to a performance orientation towards learning, while the third leads to mastery and the fourth to mastery and constructivist perspectives which they call a learning orientation.

Hattie and Timperley’s (2007) recent review of feedback literature identified that the type of feedback and the way it is given mediates effectiveness. They identified four types of feedback. The first was feedback on the task itself such as whether work is correct or incorrect. The second was feedback about the process undertaken to complete the task, trying to direct the student to use a particular approach or strategy to improve their work. The third was feedback focused on student self-regulation; here students are reminded of strategies they have been taught that they can use to improve their own work. Finally, there was feedback about the self, which Hattie and Timperley (2007) argue is often unrelated to the task; this is often in the form of praise (i.e. ‘good job’, ‘you’ve tried hard’). They posit that feedback about the self is least effective as it doesn’t actually
provide the student with any information as to how to academically improve their work. Feedback on process and self-regulation are identified as the most powerful kinds, but feedback on the task is reported as most common. Feedback relating to self-regulated learning is seen as particularly powerful (Butler & Winne, 1995).

Although academics have created models of feedback based on observing teacher actions within the classroom and empirical review of research literature, studies have not sought to understand what teacher motives and purposes underpin their uses of these practices. While there is a growing body of literature examining student understandings of feedback (i.e. J. Brown, 2007; Carnell, 2000; Higgins, Hartley, & Skelton, 2002; Lipnevich, Smith, & Barnhart, 2008; Peterson & Irving, 2008; Poulos & Mahony, 2008), teacher conceptions of feedback practices have seldom been investigated through empirical research. There are numerous models in the research literature (e.g., Askew & Lodge, 2000; Butler & Winne, 1995; Hargreaves, 2005; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006; Shute, 2008; Tunstall & Gipps, 1996), but it remains unclear how practicing teachers relate their own work to these models, if at all.

**The research problem**
The study reported in this paper was set in New Zealand and was largely exploratory. It explored the research questions:

• What conceptions of assessment are held by New Zealand secondary school teachers?
• What conceptions of feedback are held by New Zealand secondary school teachers?
• How do New Zealand secondary teachers’ conceptions of assessment and feedback relate to each other?

To address these questions, research was carried out in four New Zealand secondary schools. The results of this study are impacted upon by the New Zealand assessment culture and policy contexts. New Zealand teachers are required to follow a national curriculum, and can use observational or informal assessment methods along with a variety of teacher-administered standardised assessment tools (i.e. Assessment Tools for Teaching and Learning (asTTle), Progress and Achievement testing (PAT)) and teacher-created summative assessment pieces. The final three years of secondary school (Years 11-13) are dominated by an external qualification known as the NCEA. The NCEA is a standards based qualification comprised of both internally and externally assessed components which are designed to reflect the intrinsic nature of the standards being assessed. For example, students would have an internal assessment of a speech in English or solving a measurement problem in mathematics or conducting a laboratory experiment in science, but are externally assessed in reading comprehension of unseen text or solving a linear equation or explaining the way in which the moon affects the tidal cycle.

**Data collection and analysis**

Focus groups were chosen for data collection because they can explore perceptions, feelings, motivations, and attitudes (Kitzinger, 1995; Krueger, 1994). This data collection
technique has previously been used to explore student conceptions of assessment and feedback (e.g. Lipnevich et al., 2008; Peterson & Irving, 2008; Poulos & Mahony, 2008), making it seem appropriate for the present study.

Eleven teachers (4 mathematics, 6 English and 1 science) from four large, diverse co-educational Auckland secondary schools participated in two 90-minute focus groups, each run by a member of the research team (see Table 1 for participant demographics). These teachers came from schools with diverse decile rankings, which in New Zealand are a measure of the socio-economic status of school pupils; high decile schools are in more economically affluent areas, while low decile schools are located in areas that are more greatly affected by poverty. The teachers involved in this study were participants in the CAF (Conceptions of Assessment and Feedback) project, a larger two year study. These focus groups were run at the very beginning of the two year project to explore student and teacher conceptions of assessment and feedback and their influence on student achievement, meaning that this forum was the first time teachers had the opportunity to formally discuss these concepts. As much as possible, teachers were separated from those they worked closely with in an effort to maximize their freedom of response. Focus groups were audio recorded and transcribed verbatim.

Table 1: Participant demographics

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Subjects taught</th>
<th>Years teaching</th>
<th>Ethnicity</th>
<th>School Decile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caroline</td>
<td>Female</td>
<td>English</td>
<td>Greater than 10</td>
<td>European</td>
<td>High</td>
</tr>
<tr>
<td>Kelly</td>
<td>Female</td>
<td>ESOL</td>
<td>Greater than 10</td>
<td>European</td>
<td>Low</td>
</tr>
<tr>
<td>Harvey</td>
<td>Male</td>
<td>English</td>
<td>Greater than 10</td>
<td>European</td>
<td>Mid</td>
</tr>
<tr>
<td>Emma-Jayne</td>
<td>Female</td>
<td>English</td>
<td>Less than 5</td>
<td>Pasifika</td>
<td>Low</td>
</tr>
<tr>
<td>Rachel</td>
<td>Female</td>
<td>English</td>
<td>Less than 5</td>
<td>European</td>
<td>Mid</td>
</tr>
<tr>
<td>Annette</td>
<td>Female</td>
<td>Mathematics</td>
<td>6-10</td>
<td>European</td>
<td>Mid</td>
</tr>
<tr>
<td>Deborah</td>
<td>Female</td>
<td>Mathematics</td>
<td>Greater than 10</td>
<td>European</td>
<td>Mid</td>
</tr>
<tr>
<td>Mark</td>
<td>Male</td>
<td>English</td>
<td>Greater than 10</td>
<td>European</td>
<td>Mid</td>
</tr>
<tr>
<td>Justin</td>
<td>Male</td>
<td>Mathematics</td>
<td>6-10</td>
<td>European</td>
<td>High</td>
</tr>
<tr>
<td>Bert</td>
<td>Male</td>
<td>Science</td>
<td>Greater than 10</td>
<td>European</td>
<td>Mid</td>
</tr>
<tr>
<td>Fatima</td>
<td>Female</td>
<td>Mathematics</td>
<td>Greater than 10</td>
<td>Middle Eastern</td>
<td>Low</td>
</tr>
</tbody>
</table>

The focus groups were designed to address three key aspects of assessment and feedback: definition, purpose, and personal response. Fontana and Frey (2000) noted that engaging recalcitrant participants can be problematic. To allow all to become involved, prior to any group discussion, the teachers wrote down their definitions of assessment and feedback, their purposes, and a personal response to each on coloured Post-It ® notes/stickies. They placed these on a large piece of paper hung on the wall under relevant headings e.g., ‘Assessment Definition’ (see Peterson & Barron, 2007, for a description of this process). The stickies were collected at the end for textual analysis and triangulation with the transcripts.
These data were transcribed verbatim and labeled using Lankshear and Knobel’s (2004) method. These labels indicated which focus group the piece of data came from (represented by the numbers 1 and 2), who the speaker was (represented by the first letter of their pseudonym), and what turn this utterance was within the sequence of the focus group. For example, the passage “We still only get 31% of our kids through NCEA” was labeled 2R:036, indicating this utterance was from the second focus group, it was said by Rachel, and that it was turn 36 in the sequence of talk. Comments that participants wrote onto stickies were also labeled in this way, using the letter ‘S’ for stickie instead of a participant name as these were completed anonymously.

Coffey and Atkinson’s (1996) categorical analysis technique was used to analyse data. First, all authors independently read the focus group transcripts. After several readings, categories were allowed to ‘grow out of’ or ‘emerge’ from the data (Lankshear & Knobel, 2004). Each author independently assigned preliminary codes to these data. These codes were then compared and contrasted iteratively with codes generated by the other authors to create a set of categories that all researchers felt aptly described the data. All authors reviewed the final categorisation of the data and reached consensus on the classification.

Results

In this study, teachers discussed a range of different types of assessment and feedback. Often, they saw these as being related to different purposes. Teachers talked about three types of assessment: formative, teacher-controlled summative assessment, and external summative assessment. While teachers frequently associated formative assessment with student improvement, both types of summative assessment were most commonly associated with school accountability purposes and external summative testing was most likely to be deemed irrelevant.

Teachers talked about feedback in three ways: written/spoken interactions with students, grades and marks, and comments about behaviour and effort. Feedback was seen as having four purposes; some types of feedback were associated with multiple purposes. Feedback could tell students what they need to improve, it could encourage them to try harder, it could show parents and students the level of achievement the student has reached, or it could be irrelevant. This section explains these categories, illustrating them with sections of the transcript.

Conceptions of assessment

Teachers spoke about three main types of assessment: formative, teacher-controlled summative assessment, and external summative assessment. Formative assessment was defined in a range of ways, but was generally depicted as non-graded teacher, student, and peer interactions that showed pupils what they were doing well and where they could improve, or as the gathering of diagnostic information that could direct the teacher’s instruction. Teacher-controlled summative assessment was described as graded or marked pieces of work that could provide a reportable snapshot of student achievement. External
summative assessments were described as tests and test-like practices. These were usually considered to be high-stakes for students and, at times, the teacher; teachers reported that they had little to no control over when and how these were administered. These assessments could be mandated by the school and implemented department wide, or required by the ministry as part of nation-wide assessment.

Teachers said that these three types of assessment often had different purposes. As Fatima explained:

I think it also depends on the purpose of the assessment. If it's for a report you have to use national standards; that's how it makes it reliable. If you're using [it] for the students, it's different; one assessment doesn't serve all the purposes. (1F:041)

Of these types, they spoke most positively about formative assessment, saying this was most related to student improvement. They suggested that both kinds of summative assessment were conducted mainly for reporting and compliance purposes; these were at times seen as irrelevant. Teachers reported that while assessment should be fair and transparent, this was not always the case. While some participants saw external assessment as bringing an improved level of reliability and validity, others said it was at times inaccurate; they articulated that their teacher judgments more validly and reliably reflected student learning. The relationship between these types of assessment and their purposes can be seen in Figure 1. Within this model, dotted lines represent pathways where there weaker support or disagreement among the participants.
Teachers spoke most positively about formative assessment which they saw as being used for improvement purposes. As one teacher wrote on a stickie, ‘I LOVE formative assessment. It empowers me and students” (2S:82). Teachers saw a range of different assessment practices as being ‘formative’ as illustrated by the following discussion:

2E:016 … your formative assessment is testing that you do throughout this unit. So if I’m doing something on Whale Rider, I’ll be doing little tests throughout the unit, testing what they know. For example, I do closed activities and things like that, just testing their comprehension of the novel instead of me going right to the end of novel and not really knowing what they know so formative is just little tests you do on the way.

2R:017 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 the assumption that they don’t know how to paragraph and they do, something like that.

2C:018 Formative for me, what I do with my own students, it’s like your formative years. It’s when you’re allowed to make mistakes; you get away with much more in your formative years than you do when you’re an adult. I say this is the time when you make mistakes because I, personally, I learn through mistakes. That’s basically my only way of learning is when I’ve made a boob and I’ve been told it’s okay this time, but this time next time we won’t be so easy on you. And I find that if I tell that to the students and they know they’re getting another chance, then they’re much more ready to take it on board …

2K:019 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 or what they understand, but as you say, to actually help them to know what they need to learn next as well. Here, Emma-Jayne, Rachel, Caroline, and Kelly discussed assessment that they articulated is formative in nature. These include diagnostic assessments to determine prior knowledge, tests and activities to assess progress throughout a unit, feedback, and opportunities for students to learn from mistakes. This type of assessment is described as helping students know what they need to do to improve and assisting teachers in knowing what they need to teach to maximise student learning.

Formative assessment is described as being continuous, as Justin explained:

I guess we do assess continually by seeing where the students are at in a very casual way. It’s assessment too… There's a lot of other information between that [pieces of summative assessment] which we casually pick up on. (1J:009)

In this way, it is considered as non-threatening for students. As Caroline explained, “That’s formative assessment. I love that. It’s moving forward still. It’s not the end” (2C:140).

Teachers were less positive about teacher-controlled summative assessment. They described it in a range of ways including: a “‘convenient’ measurement for reports/exams” (2S:22); “A snapshot of progress in learning again some preset goal or
criteria” (1S:03); and “Assessment gives a level and a position on a rubric for a student” (1S:07).

Teachers indicated that grades are a main difference between formative and summative assessment. For example, Deborah explained:

One of the questions that my students said to me several times now is when I'm giving a topic test or just a test in class, “is this one that counts towards our grades?” So they are picking up that some of them are just to see where they're at and some of them are going towards their report and their parents are going to find out, but they seem to have a concern about that. (1D:011)

While the teachers articulated that students appeared unconcerned about formative assessment results, they were worried about anything contributing to their reports. This is because teachers indicated that summative results were primarily used for reporting to parents and the community; “To have grades” (2S:25) was reported as being a main purpose for these kinds of assessments. As Harvey explained:

This is more 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 (2H:079)

Teachers in this study did not talk about students as being solely responsible for these results; they reported a high level of ownership for student results as evidenced by the way Harvey used pronouns like ‘we’ and ‘our’ rather than third person pronouns like ‘they’ and ‘them’.

While some teachers were quite negative about summative assessment, with one writing “I don’t enjoy summative assessments. It feels like the end” (2S:79), several teachers discussed how they used summative results to inform future planning and teaching and indicated that some of their students also used these results to improve their learning and achievement. However, most perceived that their students did not use information from summative assessments constructively to improve their learning. For example, Caroline and Harvey discussed how students react to formal, graded assessment:

2C:089 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 maneuvering 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.

2H:090 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 for this?” and it's a final item, it’s final assessment. And kids say, “Why didn’t I get a merit? 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 credit comes from.

2C:091 You’ll only have two students doing that though won’t you?

2H:092 You won’t get a lot, but you’ll get a few.

While Caroline is clearly more negative about formal assessment than Harvey, even he articulated that most of his students see this kind of assessment as an end, with only a few using summative assessment pieces in formative ways.
Teachers articulated concerns that many formally graded assessments were confusing to students because criteria were unclear. While teacher-controlled summative assessments were school based, most teachers reported that they still had to follow department conventions, curriculum guidelines, and utilise set criteria. These criteria were seen as particularly problematic. For example, Caroline, Kelly, and Rebecca discussed how even teachers had difficult grading work consistently, making it unlikely that students could accurately self and peer assess work:

2C:058 Your descriptors that you’ve given them. which they can [use to] do their own check marking should be transparent enough, so it’s not actually a mystery why that’s an achieved and that’s a merit. I think there’s still a lot of mystery with kids.

2K:059 Yeah, I think we still have to be a lot clearer with making sure that kids understand what we mean by what we tell them. In some cases it’s not enough to say this is the descriptor.

2C:060 No it’s certainly not enough
2R:061 But we have a fight amongst teachers about what are intrusive errors because one of our teachers says, ‘If I just can’t physically make sense of the sentence, then that’s an intrusive error, but if a kid can’t manage tense, then that’s not intrusive for me because I can still understand it.” Whereas, some teachers are saying, “No, that’s an intrusive error because they do not understand tenses.”

2C:062 You’re right.

2R:063 It’s still bandied about by teachers of what descriptors mean.

2C:064 So you can see why the kids wouldn’t get it.

2R:065 Exactly, and the other thing you’ve got to take into consideration is I’ve taught 120 kids and I’ve taught this task five times or whatever and so I’m slightly more experienced at picking what an achieved is because I’ve used that descriptor X times.

Here, they explained how difficult it can be to articulate clearly to students what is expected from them according to a particular marking scheme as even teachers interpret it differently. Teachers report that this is also a reason why peer and self assessment using criteria can’t be used for reporting purposes; since students don’t have the same expert knowledge of the criteria that the teachers have, they cannot be expected to make consistent and accurate judgments.

While teachers described teacher-controlled summative assessments more negatively than their formative practices, they still preferred these to the external summative assessment like the external components of the NCEA and department-wide testing. For example, one teacher wrote on a stickie, “Much happier designing my own assessments. Marking using current NCEA methods has affected me extremely negatively (and the students too)” (1S:17). This attitude may, in part, be due to the relatively high stakes nature of much of this external testing at the secondary level. As Mark explained, “… if they fail NCEA then the future pathways for them are severely limited” (1M:138).

These kinds of assessment were seen as potentially damaging to students’ futures and self-esteem. For example, Annette explained:
I enjoy the interaction between myself and the students, watching them progress and develop within the classroom environment, watching their self-esteem grow and their views on learning, watching that improve. And then, on the flip side of it, I'm required to give them an end of term test within the department, which is set across all ability bands. And knowing that when I give them the test, that they're going to achieve a low percentage grade and by departmental pass I have to give them a percentage grade. 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 going to be internal/external and it’s almost like me and the students versus the rest. It’s coming back to what we’ve already said, we’ve got to keep an eye on the external and having that balance, but 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 thing is going to happen and we all know that we’re all going to fail (1A:099).

Annette constructed a dichotomy; it was her and her students versus the department wide test, showing that she drew a distinct divide between her personal practices and those she was required to do by external groups. Once again, she took considerable ownership of her students’ results, using the pronoun ‘we’ and including herself in the failure.

Some teachers in the group indicated that requirements relating to external summative assessment showed a lack of faith in teacher judgments, while others saw this as an important reliability check. For example, Annette and Justin, discussing this issue, said:

1A:054 I think the validity and reliability comes down to teacher assessment versus results by test and it's the internal, external view on results if you like, so even when we come to reporting to parents 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 because 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 or it can throw out, “that's a pass” or “that's a fail”. That's where politically we've got so hung up on “Is it a pass; is it a fail?” rather than the student is actually here. But I think we've got to start to hold onto our professional judgments.

1J:055 So for that you do need the reliability of staff.

1A:056 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.

While Annette clearly articulated that teacher judgments are more reliable and accurate than school-wide test data, Justin pointed out that staff judgments are not always consistent, intimating that he holds a different point of view on this issue than Annette.

The generally negative viewpoint towards external summative assessment led some teachers to see the results of these assessments as being irrelevant. As one teacher wrote on a stickie, “Assessment should be fair, transparent. Irrelevant if too external” (1S:12).

Some teachers questioned the reliability of one-off external assessments, saying:

1D:045 We were also talking about if it were reliable for just that day. There's all sorts of
other factors that affect students. Did they have a fight before they came? Did they have breakfast? Are they having a really good day that they’re engaged? So is assessment really only at that point there? We know our students all year and they get a bummer of a question and they fail, when we know that they’re actually capable, but at that day, at that time they didn't manage to come through.

1F:046 And something [that] actually affects their results is that in every single assessment at national level, the language is tested, not the knowledge of the subject as such. We have many students who [don’t] understand the question. Once we explain it to them, they can come up with the right and logical answer, but they fail in the exam because they didn't understand the language of the exam.

1A:047 But they could do the mathematics, but they don't understand the question.

1F:048 Absolutely.

Here, Daniel, Fatima, and Annette noted that multiple factors can affect the reliability of a test including the student’s personal life, and student reading/writing abilities. These kinds of factors led many of the teachers within this study to reject the results any assessments that are not ongoing.

Assessment was also seen as being irrelevant if it was not acted upon; the more disconnected the assessment was with what was being taught, the harder it was to actually use the results constructively. This was seen as putting teachers in an awkward position; on one hand they knew what students’ weaknesses were, but, on the other hand, there was a curriculum they were required to teach:

2K:133 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.

2C:134 And I could do that but I’m not feeling very happy about that.

2K:135 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.

Here, Kelly and Caroline pointed out that teachers are often under pressure from competing forces; some feel they have to ignore the assessment information because of school and curriculum requirements related to the quantity of information they need to cover.

To summarise, teachers reported three main purposes for assessment. All teachers articulated that formative assessment was useful for the purpose of student improvement and was a positive and useful tool. While some said that teacher-controlled and external summative assessments were useful for accountability purposes, others indicated that they were, at times, irrelevant and inaccurate. Few teachers reported using summative results for formative purposes or seeing their students using their formal results to improve. While teachers said that expectations and criteria should be explicit and transparent, they indicated that this is not regularly the case and highlighted the negative
emotions their students can feel due to summative assessment. Assessment was also considered to be irrelevant if it was not acted upon. Teachers articulated that they were under competing pressures; while they wanted to use assessment data to help students improve, they also felt pressure to cover a set amount of subject content with the class because of curriculum guidelines.

**Conceptions of feedback**

Teachers described three types of feedback: diagnostic feedback about learning, grades or marks, and feedback on behaviour. While feedback about learning and behaviour could be oral or written, grades were depicted as a written form of feedback. While teacher conceptions of assessment and feedback were in some ways related, there were key differences. First, participants articulated that there was a fourth purpose for feedback, encouragement; this was not seen as being a purpose of assessment and did not directly relate to any of the other purposes of feedback. Second, teachers indicated that timing matters with feedback; while diagnostic feedback about learning was generally seen as being for improvement purposes, it was viewed as irrelevant if given along with a grade or mark because they said that students disregarded it. The relationship between conceptions of feedback is shown in Figure 2.

![Figure 2 - Teacher conceptions of feedback](image)

Teachers talked primarily about giving feedback about learning. They described a range of different practices including: “A tick in the margin is feedback. Likewise a question mark etc or a smiley face” (2S:56), and “Could include feed forward” (2S:59). Justin and Daniel described this process more thoroughly, explaining:
1J:027 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 talking to them would be ideal but more often it is just a note on perhaps an assessment they’ve done or some work that they’ve done just pointing out what they’ve got wrong and how they can get that bit correct.

1D:028 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, you know positive and negative ...... 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” 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 appears with my teaching, “Everybody has understood this, but there's these little bits as a whole or as individuals that you need to improve on,” so that's more just as a verbal general impression.

This mixture of oral and written feedback is seen as being important in that it lets students know what they need to work on in order to improve. As Bert explained, “I believe that kids do progress because we give them specific feedback and so we break down the skills” (1B:023).

While teachers were generally negative about summative assessment, several acknowledged that grades and marks are a form of feedback. For example, Harvey explained, when discussing how the stickies people had written about feedback could best be organised:

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 [the other side] is the diagnostic and informative, well diagnostic and figuring out what to do next side of things I suppose. See where the gaps are, that one goes on this side. (2H:068)

Here, he differentiated between diagnostic feedback and grades, but articulated that both are types of feedback. The purpose of grades as feedback was for reporting purposes. As one teacher explained, “Feedback validates student progress up to a certain point. Part of the discourse of a classroom” (1S:20).

Teachers also talked about feedback given to improve students’ self-esteem and to acknowledge and encourage positive classroom behaviour. As Mark explained, “We give feedback about behaviour: “Oh, that's great the way you put your hand up or whatever,” but we also give feedback about learning” (1M:014). Feedback about behaviour and effort was “To encourage students” (2S:74)

Giving kids encouraging feedback was seen as important when dealing with low achieving students. As Annette explained:

I'm thinking a bit about the grouping taken for this project. It's a Year 10 group and I would say they're working at about Level 4 in mathematics and the main thing at this point that I'm commenting on in their books is about their attitude and their effort and trying to engage them. I do correct them subject specific, but because their self-esteem is somewhat low, at this point the major thing that I'm
writing in their books is how well they've done, how good it looked, where in the classroom they were able to go up to the board and thanks for helping so and so out. (1A:031)

Annette’s description is typical of descriptions of this kind of feedback; she explained how she overemphasized positive aspects of their work to ‘try to engage them’. This kind of feedback is seen as promoting positive affective feelings in students and making them more likely to engage in learning.

These three types of feedback were seen as having four different purposes: improvement, reporting, irrelevance, and encouragement. All four are described in the conversation below:

2K:111 I do very little assessment by written work. Most of my feedback would be oral feedback and when I’m doing my one to one work I assess the reading by doing a sort of a running record every time you read to me. But that’s always informal, but at the end of the week, I always put it 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, how well they’re doing and how soon they’ll be at the top sort of thing and achieve their aim. It’s always discussion, so I suppose that’s giving a mark and a comment.

2C:112 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. It’s like, “Hey you did a really [good job], you’re so nearly there.” I always use their name, more for me, trying to pick the kids’ sort of self-esteem up and I don’t know whether it works or not. But I do find it hard; I would just find it hard to put a not achieved, even though I know that’s probably the most efficient use of my time to do that and say even as I’m handing it back to say, “Hey, you nearly did,” but it’s a politeness thing.

2K:113 I know it’s time consuming and stuff, but I think even somebody who’s got an excellent 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.

2H:114 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 their way through the comment to get to the grade. But I put, “Very good beginning; very strong opening. There are some intrusive errors in paragraph nine blah, blah, blah. This would get a merit if you had tidied up your language.” And that will be written in it and that kind of, 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, Might not at junior level; it might not be an issue. But at senior level, it is an issue; they need to know that they are within the ball park of getting a such and such or a so and so.

This discussion shows examples of all four purposes of feedback. Teachers talked about feedback being for improvement, like when Kelly explained why it is important to tell students what they did well so they can continue to achieve at that level. Teachers
disagreed over whether it was appropriate to include grades along with comments; when this was done, it was for reporting purposes so kids knew what to expect on their final piece. While Kelly and Harvey said this was part of their normal practice, Caroline disagreed because, as she articulated earlier:

...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, I say, “No, there’s not going to be a grade; you’re just getting comments in your formative assessment,” 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. (2C:100)

This is likely to be why she noted that while she always writes comments on summative assessment to help protect students’ self esteem, this is probably ‘pointless’ and irrelevant; most teachers agreed that feedback on summative pieces was largely irrelevant as students ignored it. This is an example of how timing was seen as crucial in the delivery of diagnostic comments; if students did not receive them prior to summative assessment, the comments were perceived as irrelevant.

Feedback was also seen as used for reporting purposes. As Harvey mentioned, by burying a ‘ballpark’ grade in his comment, he was clearly signaling to the student (and other stakeholders like parents) what he/she could expect to achieve given the current standard of the work. As one teacher wrote on a stickie, feedback is “To keep parents and students informed of progress and needs” (2S:69), showing that, at times, teacher feedback has a wider audience than just the students. Encouragement was also a purpose, explaining why Caroline chose to write comments on her student work, even though she stated it was pointless.

To summarise, the teachers in this study described three types of feedback: feedback about learning, grades or marks, and encouraging feedback. Feedback about learning was seen as leading to student improvement unless it was accompanied by grades or marks; in this case, most teachers described it as being irrelevant as students ignored the comments. Grades and marks were seen as being a performance indicator for students and useful for reporting purposes. Teachers also noted that especially with low performing students, encouraging feedback about behaviour and effort were important. These comments were not achievement based and teachers articulated that they hoped that these would make students more engaged and that this engagement would help them learn.

Conclusions

This study indicates that teacher conceptions of assessment and feedback are closely related, but that there are also distinctions. Teachers indicated that the purpose of assessment and feedback can be seen in one of three main ways. They:

• improve student learning,
• are used to report student performance to stakeholders and comply with school and ministry regulations
• are detrimental or irrelevant to student learning.
With feedback, encouragement is seen as an additional purpose. A model of the relationship between conceptions of assessment and feedback is shown below in Figure 3.

Contrary to expectations, teachers did not indicate that assessment should make students accountable for their learning, as predicted by G. Brown’s (2002, 2004, 2006, 2007) research. Instead, teachers reported taking a high level of personal ownership for their students’ successes and failures when it came to assessment, frequently using inclusive personal pronouns like “we,” “us,” and “our” when describing student results. They also articulated that reporting was an important, although at times unpleasant, purpose of assessment indicating that most did believe that schools and teachers should be accountable to parents, students, and the community for student learning. However, teachers reported disliking the current standardized department and nation-wide tests used to determine student results, citing that these were unfair to certain student groups and that they were often invalid and unreliable.

As this study also asked teachers for personal reactions to these concepts, their affective response was also examined. While teachers were universally positive about feedback and its purposes (even with feedback they said was ignored, they noted that it made them feel better as teachers when they wrote these comments), teachers had more divergent views about assessment purposes. While all cited extremely positive feelings about formative assessment, teacher-controlled summative assessment was viewed in slightly
negative terms, and external summative in extremely negative terms. Teachers said that they wanted assessment to be a positive experience for their students, something they found to be at odds with high-stakes, top-down evaluation of students like that which occurred through external summative assessment. Teachers expressed that there was a tension between wanting to do formative assessment and being forced to use summative for reporting and compliance purposes.

Another important finding was that most teachers in this study did seem to view formative and summative assessment as dichotomous, something authors like Hargreaves (2005) note as unproductive. These data indicate that more work is 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 of reporting and achieving compliance.

Also, when discussing feedback, most teachers indicated that feedback about behaviour and effort was for the positive purpose of encouraging students to engage. This belief is at odds with what Hattie and Timperley (2007) report in their review of feedback literature. Hattie and Timperley (2007) note that feedback about the self is “…rarely converted into more engagement, commitment to the learning goals, enhanced self-efficacy, or understanding of the task” (p. 96). This paradox is an important one that merits further investigation as at present it is unclear why teachers believe this kind of feedback is effective even though research suggests otherwise.

This was a very small scale study, but it does provide interesting and useful data to help understand teacher conceptions of assessment and feedback and how these might relate together. While only 11 teachers were used, these participants came from diverse schools and had a range of teaching experience, making it a diverse sample. The model of how teacher conceptions of assessment and feedback relate together is a theoretical one based on this empirical data set and relying on self-reported data. Future studies should investigate these relationships using a range of qualitative and quantitative data to test the model.

References


