

New Zealand teachers' conceptions of the purpose of assessment:
Phenomenographic analyses of teachers' thinking

Lois R. Harris & Gavin T. L. Brown
The 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
a University of Auckland Faculty of Education Faculty Research Development Fund
grant (#3609273/9215) to second author. The cooperation and participation of the
teachers and schools in the Measuring Teachers' Assessment Practices (MTAP)
project is appreciated.

New Zealand teachers' conceptions of the purpose of assessment: Phenomenographic analyses of teachers' thinking.

By Lois R. Harris and Gavin T. L. Brown, University of Auckland

Abstract

Introduction

Assessment is an extremely important part of the teaching and learning cycle. As Boud (1995) noted:

Students can, with difficulty, escape from the effects of poor teaching, they cannot (by definition if they want to graduate) escape the effects of poor assessment. Assessment acts as a mechanism to control students that is far more pervasive and insidious than most staff would be prepared to acknowledge. (p 35)

Entwistle's (2000) research showed that many students adopt different approaches to learning based on the way they perceive they are going to be assessed. Because of assessment's power to shape student approaches to learning and to facilitate or block desired pathways, it is important to more fully understand how teachers' beliefs and conceptions underpin particular assessment practices. Furthermore, teacher conceptions are important since these influence their own practices (Clark & Peterson, 1986; Pajares, 1992; Thompson, 1992) which, in turn, affect student outcomes (Muijs, 2006; Muijs & Reynolds, 2005).

Debates over best practice for assessing students stem partly from divergent ideas about the purposes of assessment. Researchers (e.g. Brown, 2004; Heaton, 1975; Shohamy, 2001; Torrance & Pryor, 1998; Warren & Nisbet, 1999; Webb, 1992) have suggested that stakeholders hold four major conceptions about the purposes for assessment:

- It improves teaching and learning (Improvement).
- It makes students accountable for learning, partly through issuing certificates and credentials (Student accountability)
- It demonstrates the quality and accountability of schools and teachers (School accountability)
- It should be rejected because it is invalid, irrelevant, and negatively affects teachers, students, curriculum, and teaching (Irrelevance)

The premise of the improvement conception is that assessment should improve students' own learning and the quality of teaching (Crooks, 1988; Black & Wiliam, 1998; Black, Harrison, Lee, Marshall, & Wiliam, 2003). This improvement has two important caveats: (a) assessment must describe or diagnose the nature of student performance and (b) the information must be a valid, reliable, and accurate description of student performance. A range of techniques, including informal teacher-based intuitive judgment and formal assessment tools, are used to identify the content and processes of student learning with the explicit goal of improving the quality and accuracy of instruction and/or enabling students to improve their own learning.

The premise of the second conception of assessment is that assessment can and should hold students individually accountable for their learning. Practices aligning with this conception include: assigning grades or scores to student work, checking student performance against criteria, and awarding certificates or qualifications based on performance. In New Zealand primary schools, the use of assessment for student accountability focuses much more on determining whether students have met various curriculum objectives (Hill, 2000), the criteria for a given curriculum level (Dixon, 1999), or merit placement in a certain learning group within a class. At a secondary level, there are the various qualification assessments for school completion or entry selection to higher levels of educational opportunity. This external validation occurs primarily during the final three years of New Zealand schooling; consequences for individuals include: retention, graduation, and tracking or streaming (Guthrie, 2002).

A third conception is that assessment can be used to account for a teacher's, a school's, or a system's use of society's resources (Firestone, Schorr, & Monfils, 2004). Here, assessment is for the purpose of demonstrating publicly that teachers or schools are or are not effective (Butterfield, Williams, & Marr 1999; Mehrens & Lehmann, 1984; Smith, Heinecke, & Noble, 1999). It is used to impose consequences for schools or teachers that achieve or do not reach required standards (Firestone et. al., 2004; Guthrie, 2002).

The premise of the final conception is that assessment, usually understood as a formal evaluation of student performance, has no legitimate place within teaching and learning. This conception is underpinned by the belief that teachers' knowledge of students is based on long relationships and their expert understanding of curriculum and pedagogy, making it unnecessary to assess beyond the intuitive in-the-head process that occurs automatically as teachers interact with students (i.e., Airasian's [1997] 'sizing up'). Assessment may be rejected also because of its pernicious effects on teacher autonomy and professionalism and its distractive power from the real purpose of teaching (i.e., student learning) (Dixon, 1999). It may also be that the degree of inaccuracy (e.g., standard error of measurement) published with any formal assessment contributes to teachers' conception of assessment as irrelevant. Some teachers also reject assessment because they feel it is inherently unfair to certain types of students, it causes children to become unnecessarily anxious, and that poor performance may damage student self-esteem (Harris, Irving, & Peterson, 2008).

During the last decade, there has been an increased focus on formative assessment practices, under the rubric 'assessment for learning' (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Black & Wiliam, 1998; Clarke, 2005). These recommendations align with the conception that the purpose of assessment is improvement. Formative practices are often contrasted with more formal and summative practices such as standardised tests which are commonly associated with both accountability and irrelevance conceptions. Over-emphasis on accountability has been shown to have negative consequences for teachers, teaching, curriculum, and learning (Firestone, Schorr, & Monfils, 2004; Hamilton, 2003; Jones, 2001).

Using the Teachers' Conceptions of Assessment (TCoA) survey instrument, Brown and colleagues have investigated New Zealand (Brown, 2004, 2006b, 2007), Australian (Brown & Lake, 2006), and Hong Kong (Brown, Kennedy, Fok, Chan, & Yu, in press) teachers' levels of agreement with the improvement, student

accountability, school accountability, and irrelevance conceptions. Across countries and sectors, teachers consistently had positive agreement means for both improvement and student accountability conceptions and had low means for school accountability and irrelevance conceptions. By inspection of the inter-correlations among the four factors, it could be seen Australian and New Zealand teachers conceived student accountability as being irrelevant, while school accountability was correlated with improvement. In contrast, the Hong Kong teachers had strong positive correlations between improvement and student accountability. There were moderate differences between primary and secondary teachers in New Zealand and Queensland especially around the importance of student accountability. Together these studies support the claim that there are strong contextual factors in teachers' thinking about assessment.

While the TCoA inventory research has provided robust data relating to teachers' conceptions of assessment, it has some limitations. First, it measure responses to just four conceptions of assessment (i.e., improvement, student accountability, school accountability, and irrelevance). While validating evidence was found during the development of the inventory (Brown, 2002), there may yet be significant purposes that are not captured by the inventory. Further, the development depended on analysis by one researcher and few independent studies concerning the validity of the inventory have been conducted. Additionally, while the structure of teachers' conceptions can be detected, the anonymous survey method is unable to ascertain reasons teachers might have for their responses. Hence, in the study reported in this paper, an independent researcher collected data through an interview and analysed the results interpretively to establish the range and structure of conceptions teachers had about assessment.

This study adopted a qualitative approach to investigate teacher conceptions of assessment and sought to answer the question:

What qualitatively different conceptions of assessment are held by New Zealand teachers of years 5-10?

As this question deals with variation in conceptions, a phenomenographic approach (Marton, 1981, 1986; Marton & Booth, 1997) was selected to investigate this question.

Data collection and analysis

The study reported in this paper is part of the Measuring Teachers' Assessment Practices (MTAP) project at The University of Auckland. All primary, intermediate, and secondary schools in the greater Auckland region were initially invited to participate in the project. In New Zealand, primary schools include year levels 1-6 (with full primary schools catering for years 1-8), intermediate schools house years 7 and 8, while high schools include year levels 9 to 13. Over thirty schools agreed to give access to teachers of mathematics and/or English working with students in Years 5 to 10. All such teachers were asked to complete Brown's COA-III abridged questionnaire (Brown, 2006a). In all, 169 teachers participated in the questionnaire with over 100 agreeing to be interviewed. The second author analysed these questionnaire data and selected 26 teachers out of the ones willing to be interviewed who exhibited noticeably different profiles in their conceptions of assessment. Table 1 describes demographic characteristics of the participants who were interviewed.

Socio-economic status is indicated by school decile which is a government index of the status attending each school, with 1 being the lowest and 10 being the highest.

The first author then conducted semi-structured interviews with these teachers without being aware of each participant's conceptions profile. All data were transcribed verbatim and each utterance was labeled as per the method described in Lankshear and Knobel (2004). For example, in the label L1:032, the L indicated this was data from participant Lisa; as there were 26 participants, each is represented by a letter of the alphabet. The 1 represented that this was from her first interview (as part of the MTAP project design, several participants were interviewed again in later studies, but these data are not presented in this paper). The 32 indicated that this was the 32nd utterance within the sequence of the interview.

The first step in the analytical process was *bracketing* preconceived ideas (Ashworth & Lucas, 1998, 2000; Marton, 1994). Several steps were used to minimize researcher subjectivity. First, during data collection and preliminary analysis, no academic literature about assessment was read to minimize the risk of manipulating data to emulate previous findings. Second, all codes used when classifying participant data were developed from the transcripts using participant words. Third, participant data were systematically and iteratively compared and contrasted with other participant data; at no point were pre-existing categories utilized.

Analysis followed Marton's (1986) procedures and was comprised of two main steps: creating *categories of description* and ordering the *outcome space*. Categories of description were formed by grouping together similar understandings, each representing a qualitatively different way of experiencing the phenomenon. After several readings of the data, utterances related to the research questions were selected and marked using three indicators:

1. Frequency - how often an idea was articulated
2. Position - where the statement was situated; often the most significant elements were found in an answer's beginning
3. Pregnancy - when participants explicitly emphasized that certain aspects were more important than others. (Sjostrom & Dahlgren, 2002)

Once data relating to conceptions were identified, passages were analyzed and interpreted within their contexts before being removed to create *pools of meaning*. Data were compared and contrasted, leading to movement between pools. Borderline cases were examined and criteria for each pool made explicit. Once the system was stable, data in the pools were abstracted into categories of description (Marton, 1981).

Table 1- Participant characteristics

Pseudonym	Sex	Level taught	School Decile	Ethnic Group	Years of teaching experience
Kuval	Male	Primary	1	Asian Indian	More than 10
Tom	Male	Primary	1	New Zealand European	Between 2 and 5
Lisa	Female	Primary	2	New Zealand European	Less than 2
Xavier	Male	Primary	4	New Zealand European	Less than 2
Oliver	Male	Primary	5	New Zealand European	Between 2 and 5
Grace	Female	Primary	6	New Zealand /Pasifika	More than 10
Isabel	Female	Primary	6	New Zealand European	Between 2 and 5
Madison	Female	Primary	10	New Zealand European	More than 10
Pearl	Female	Primary	10	New Zealand European	Between 2 and 5
Wynona	Female	Full primary (intermediate)	2	New Zealand European	More than 10
Chelsea	Female	Full primary (intermediate)	10	New Zealand European	More than 10
Danielle	Female	Intermediate	3	Indian	More than 10
Rebecca	Female	Intermediate	3	South African	More than 10
Yvonne	Female	Intermediate	3	New Zealand Italian	More than 10
Alicia	Female	Intermediate	4	New Zealand African	More than 10
Henry	Male	Intermediate	4	New Zealand European	More than 10
Ju-long	Male	Intermediate	4	New Zealand Chinese	Between 6 and 10
Ursula	Female	Intermediate	4	New Zealand European	More than 10
Emma	Female	High school (English)	2	New Zealand European	Less than 2
Nicole	Female	High school (English)	2	New Zealand NZ Maori: Tainui (Iwi)	Between 6 and 10
Sylvia	Male	High school (English)	6	New Zealand European	More than 10
Quinn	Male	High school (mathematics)	2	New Zealand European	More than 10
Vince	Male	High school (mathematics)	2	New Zealand European	Between 6 and 10
Bimala	Female	High school (Mathematics)	3	Asian Indian	More than 10
Zac	Male	High school (mathematics)	3	Asian-Fiji Indian	More than 10
Fred	Male	High school (mathematics)	8	New Zealand European	Between 2 and 5

The categories of description were then organized into an outcome space representing all the possible ways the population under study experienced the phenomenon (Åkerlind, 2002). The creation of the outcome space relied on a balance between

allowing the structure to emerge from the data and relying on the professional judgments of the researcher (Walsh, 2000). Bracketing no longer occurred during this step as creating an outcome space:

... allows the phenomenographic researcher not only to list people's conceptions in the form of categories of description but also the researcher's interpretation of the relationship between them. (Yu, 2003, p. 5)

In this step, categories established from the data were compared with each other and with academic literature to create a 'logical' model of how the categories relate to each other. Relationships between categories were primarily defined by similarities and differences (Marton & Saljo, 1997) and categories were ordered by complexity (Marton, 1994). In this study, pregnancy (Sjostrom and Dahlgren, 2002) was relied on heavily to order conceptions; categories placed higher in the hierarchy often included participant statements critiquing lower ones. While interjudge reliability was not used in this study (see Sandburg, 1997 for reasons why this process is not generally used in phenomenographic work), the categories and final outcome space were thoroughly interrogated by the second author of this paper and were found to be descriptive of the data set.

Results

This study found that teachers hold complex conceptions of the purpose of assessment. They reported assessing formally and informally both student academic progress and achievement as well as student behaviour and socialisation. As they employed a wide range of assessment practices with their students, many appeared to compartmentalise the assessments and use different types for different purposes. Teachers conceived of seven major purposes of assessment which were abstracted into the following categories of description:

1. Compliance
2. Reporting to the board and outside entities
3. Reporting to parents
4. Extrinsically motivating students
5. Facilitating group instruction
6. Teacher use for personalising teaching and learning
7. Joint teacher and student use for personalising teaching and learning

Each of these is described in this section and illustrated with quotations; the relationship between these categories is shown in Figure 1. While all teachers commented in some way positively or negatively about the first five categories, teachers were divided over whether assessment was for teacher use or joint teacher-student use, causing the model to branch into two directions at the top.

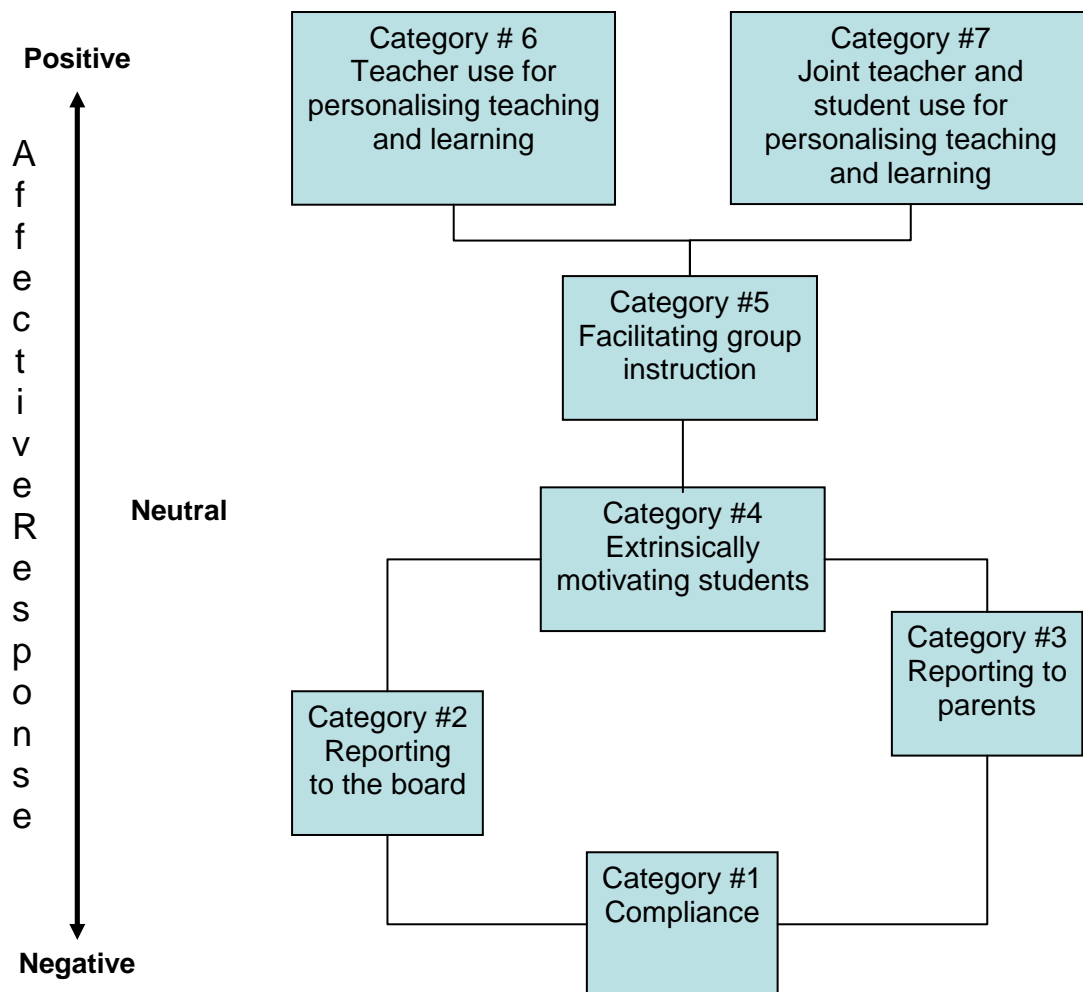


Figure 1- Teacher conceptions of the purpose of assessment: Outcome space

Category #1- Compliance

For some teachers, the purpose of using specific assessment types was to comply with school or government policy. Teachers articulated that there were some assessments that they were required to use with their students because of ministry mandates or school-wide directives. Teachers most frequently talked about standardised testing within this category, but other school based practices that clashed with the teachers' personal beliefs about effective and useful assessment were also mentioned. These practices were rejected because they were seen as irrelevant, inaccurate, or bad for pupils. Teachers were especially worried about how these kinds of assessments might negatively affect their students and expressed that they wanted to protect pupils from inappropriate evaluations of their progress and achievement.

Teachers indicated that there are assessments they are required to conduct that clash with their personal beliefs about useful and effective assessment. For example, Tom explained:

... One of the purposes is to conform with what is required under the legislation. And yeah, that's, that is why I think a lot of it is done within schools. I'm not just talking about this school. (T1:086)

Alicia commented that while teachers can try to align mandated practices with their own preferred ones, they must meet these requirements saying, “Well, they pay my salary, so whether I agree with it or not, I have to do it” (A1:078).

Teachers reported that individual schools mandate assessment practices as well as the Ministry of Education. As Fred explained:

Yeah, our school is pretty big on asTTle... Teachers get annoyed about it because it's extra work, takes them out of teaching time, and you don't see where it's going, and all that sort of stuff, but the school is really pushing it so, um, what we're doing with it yet, I'm not too sure. (F1:032)

This statement reflected teachers' stated responses to clashes between personal ideology and mandates; they comply because they feel they have to. Here, Fred admitted that he lacked knowledge about how this assessment should be used, saying he was 'not too sure.' While teachers have access to lots of assessment data about their students, this does not mean they have the knowledge and skills, or possibly even the power, to utilise the information accurately and effectively.

Some assessment data were also considered wasted because, due to over-assessment, teachers lacked time to interpret and use it properly. For example, Tom explained:

If you have a system set up whereby you are meeting the requirements as to what you have to do, then continued testing doesn't lead to an improvement in educational results. If you're on a diet, you don't lose any more weight if you weigh yourself 20 times a day than if you weigh yourself once a week and it's the same with assessment. (T1:032)

He noted that continued use of a variety of assessments is not going to positively affect student outcomes unless they are correctly interpreted and utilised by teachers, something difficult to do when the volume of assessment becomes high. Teachers said they were only able to use a small percentage of it properly for instructional purposes; much was said to be used superficially or not at all.

Many teachers worried about the accuracy of standardised testing, citing reliability issues and perceived barriers to student success; these concerns caused some to say they generally disregard this type of assessment. Teachers cited that some of their students get answers wrong because of low English proficiency or literacy levels, not lack of subject knowledge. For example, Lisa said:

Like the PAT test, the questions are quite wordy and their English is not so great, so they might know the answer, but the question confuses them and they get it all wrong. (L1:022)

Oliver explained, “I mean asTTle, to be quite honest, is a waste of time for half of my year 5 because they just look at a bit of paper with too many words and go sorry” (O1:134). These types of concerns were especially common among teachers at lower decile schools where students were perceived as having low literacy skills.

However, teachers also cited a range of other reasons that jeopardised the reliability of test results. For example, Erin explained:

[Testing is] Still very dependent on, sort of, the outside variables. You know, the weather, it sounds ridiculous but that affects, and you know, the weather, other kids, and things at home, cheating, all the rest of it. (E1:152, 154)

Many teachers described any assessment that was not ongoing and iterative as being susceptible to inaccuracy. As Tom mentioned, “Any assessment, it depends on the day

in which it's given, on the environment which the student is coming from. It may be that the student can do the same assessment a week apart and get totally different results" (T1:114). His comment reflected teacher concern about the reliability of any assessment which requires a student to perform at a specific point in time. Marking of complex student data (e.g. multi-step mathematical problem solving, writing) was also seen as inconsistently marked, with teachers like Fred noting, "... there's not often consistency among markers..."

Teachers also discussed the negative affective feelings that evaluative assessments, especially tests, can create. For example, Xavier describing testing in his own class said:

You see the looks on their faces, especially those ones that they know they're going to struggle... it almost broke my heart after she called herself dumb... a kid thinks of themselves as dumb, you're losing the battle already. So yeah. So whenever you get a grade, it's never great because a kid can see, I only got 12 percent and everyone else got 60. Who's that really helping? (X1:166)

Xavier's also noted that certain students 'know they're going to struggle'; some teachers accepted that particular pupils were bad at testing and that this was a fixed quality, making this mode of assessment perpetually unfair to these students..

While teachers reported that certain types of assessment, especially whole-school standardised tests, as irrelevant, inaccurate, or negative, they cited that they used them to comply with Education Ministry or school requirements. Most described minimal usage of these data, meaning for these teachers, such practices are assessment for the sake of assessment. Teachers reported being caught between requirements to report this kind of evaluative data and their professional judgment that these results were inaccurate and/or will negatively affect and categorise their students. However, none cited resisting these required practices despite their personal objections, meaning that in general, teachers do comply with assessment mandates regardless of personal beliefs.

Category #2- Reporting to the board and outside entities

Teachers cited that some assessments were completed for the purpose of reporting information to their school board and external monitoring organisations like ERO (Education Review Office). This kind of reporting was seen as based primarily on standardised achievement data and was generally accepted as necessary despite the qualms some had about the validity and reliability of standardised testing, discussed in the previous category. They reported that these data were used to measure the school's academic progress and compare the school's results with other schools and national benchmarks; while they disliked holding individual students accountable for their scores, when aggregated, they appeared more comfortable with uses of these data. Most of the teachers in this sample, with the exception of those holding management positions, indicated that they played little or no role in the interpretation of this data.

Most teachers in this sample agreed that schools needed to provide some form of data about the learning of their students to external sources. For example, Vince explained:

I think it's important that schools are answerable to what is happening. And if everything's standardised, then you can say you're performing well, you're

not. Why not? And if you can answer that, there's no problem. If you can't answer that, then there is a problem. (V1:190)

This kind of response was typical and reflects the low stakes assessment system in place in New Zealand schools; as long as schools can justify their results and show evidence that systems are in place to improve the learning of groups not reaching the benchmarks, they are not penalised for low achievement. While such assessments may not be beneficial for the individual students who complete them, their aggregated data is seen as useful for schools and society.

Teachers reported that this kind of whole school data was used to make decisions about resources and school-wide teaching practices. For example, Ursula explained:

We want to know where we came from and where we're going to. Um, you want to know, particularly in a school like this, what achievement looks like in this school? And where do you target money? Where do you target teacher resources? Where do you target book resources? ... Are the children actually progressing or are they standing still? And you know, are we getting closer to norm? (U1:078)

As Ju-long explained, "It's got to be [done], so that the school has got something to compare with the national mean" (J1:118).

Dealing with this school-wide data was seen as the responsibility of the school's management and was not something that general teachers reported being involved with. For example, Bimala noted:

The data [from school-wide assessment and previous school years], we don't keep the data; we just keep the record of our classes... I don't know much about the data, so who keeps it. Maybe one of the senior members, maybe one of the deputies, the level of those people; I don't know who is doing that. But someone is doing it; that's what I know (B1:076).

However, all teachers discussed more involvement with the other main means of school accountability, ERO [Education Review Office] inspection visits. However, teachers like Danielle highlighted that the heightened level of accountability only occurred during years when the school was actually being visited. She explained, "Some teachers would also say we're not having ERO this year, so it's okay," (D1:116), reflecting her perception that certain teachers behave differently when they know they're being reviewed.

While reporting data to their board and the ministry was considered acceptable, teachers perceived that data will be unethically manipulated if too much accountability was required. For example, Madison explained:

... There's lies, damn lies, and then there's statistics... if it becomes a situation where we are that accountable, schools will fib. They will tell lies or they will twist it so that it shows, they will say to certain children you are not sitting this test because you're going to bring our results down. (M1:104)

Madison indicated that if further accountability was demanded, she and other teachers would 'teach to the test' and that this would be detrimental to student enjoyment of learning and the delivery of a well-rounded curriculum. Even under the current assessment regime, teachers voiced concerns about the manipulation of whole school assessment data. Henry noted, "... I think that the information can be used and released as suits ... I think schools release what information they want to, and they don't want to release negative [data]... (H1:084). This was seen as an issue especially

at a secondary level with NCEA results often being published by the media, which they indicated was an unfair comparison between schools.

A tension existed within the data between external moderation and teacher professionalism. For example, Tom explained:

Because even without the assessment, you know, the necessary assessment and reporting to ERO and so on, I still believe those things in the main would be done, in the vast majority of cases. (T1:104)

Tom followed up this statement by explaining that he believed teachers were involved in education for altruistic reasons and that, at times, the push for standardized evaluation was shifting attention away from the needs of children as individuals and devaluing teacher judgments and professionalism.

While teachers often described standardised testing as unreliable and invalid, they conceded that it was the only way to gain a global picture of a school's current levels of achievement in comparison with other schools and national norms. This whole school analysis was not something that rank and file teachers report being involved in; however, they note that ERO visits more heavily involved all members of the school. While teachers acknowledged that it was important for them to be accountable to their school board and the ministry, they stressed that a delicate balance must be maintained to avoid situations where teachers and schools manipulated data and delivered narrow test-related instruction to students.

Category #3- Reporting to parents

While this category also deals with reporting, teachers viewed parents and external organisation as distinctly different stakeholders with differing claims on assessment information. Although teachers expressed qualms about the types of data they perceived that parents wanted, overall, they discussed reporting to parents as a useful purpose of assessment. Teachers described reporting assessment data to parents in a range of ways including formal reports, conferences, telephone conversations, and informal discussions; primary and intermediate teachers discussed utilising a wider range of assessment types than secondary teachers. Teachers in high decile schools reported more pressure from parents during the reporting process than lower-decile teachers.

Teachers consistently said that parents were an important educational stakeholder and indicated that reporting to them was a necessary part of the assessment cycle. For example, Xavier stated, "I think that's obvious. There needs to be some sort of reporting because parents want to know and that's important to them and parents need that buy-in" (X1:094). Teachers spoke of purposely trying to contact parents and get them involved in their students' education.

While teachers were happy to report to parents, some had concerns about the kind of data they perceived that parents wanted. As Ursula explained:

... for parents, as a teacher what I'm trying to do is inform them of what their child can do and how we can work together to move the child on, but for parents ... they want to know is my child average, above average, or below average. Sort of a bit of a clash there ... There's a little bit of a difference between "This is what your child can do and this is what they need to work on

next” and “This is the trend,” the parent saying, “Is my kid average or above average or below average?” That’s about giving marks. And that’s about competition for the whole group. ... (U1:078, 80)

Teachers perceived that parents found this numerical data more concrete and tangible than qualitative evaluations of their child’s learning.

Teachers voiced concerns that numerical data allowed parents to make comparisons that weren’t positive or helpful to student learning. For example, Madison explained:

What in the end, what in the end is it actually going to say? Your child was third in the class, or your child was 23rd in the class. I mean how is that really going to help you? You can see there whether your child is reading at an age that is acceptable. You know, if they’re 11 and they’re reading at 5 to 6, it’s probably not acceptable and you probably didn’t need someone to say they were last in the class for that to dawn on you. ... It [data like reading ages] does allow for that, but it doesn’t slot your child into a winning or losing. (M1:098)

Teachers seemed especially concerned about the effect of this data on lower achieving children, stating that they preferred to give data like ‘reading ages’ coupled with teacher judgments of progress rather than comparative data. Teachers also noted that not all parents wanted data about assessment as some were more concerned their child’s behaviour and attitude.

Primary and intermediate teachers discussed reporting different kinds of data to parents than their high school counterparts, using a much wider range of practices when making judgments about student learning. For example, Kuval, a primary teacher explained, “we have our parent interviews ... so we report to them and so we use all the assessments that we did, whether it’s formal or not” (K1:082). This contrasted with high school teachers who said they relied almost entirely on formal practices because, as Vince explained, “you’ve got to be able to back everything up” (V1:100).

While generally teachers were positive about reporting to parents, several did express concerns about some expectations parents had, especially parents at high decile schools. As Fred, who taught at a decile 8 school, stated:

Yeah, they’ll [parents] push the child, and the child will raise their standard or focus or whatever, but that doesn’t explain why the teachers make sure they keep telling us [new teachers], “Keep a record of all the stuff that you’ve been doing and if you’ve handed out things to help them, make sure you’ve got a record of it because when the parents do come back at you can cover your ass,” that kind of stuff. ... Obviously I’d better do that because there will be a time where parents will put it on me. (F1:124)

Extensive record keeping was seen as one way of helping teachers defend their judgments and avoid blame for poor student performance.

While teachers spoke positively about reporting to parents, concerns were echoed over the types of data they perceived that parents wanted. While primary and intermediate teachers reported using a range of assessment data for reporting purposes, secondary teachers said they relied on formal assessment as it was perceived as more credible and defensible. Teachers in secondary and higher decile schools reported described their school’s parents as expecting a high level of

accountability and described more concerns relating to being blamed for poor student achievement.

Category #4- Extrinsically motivating students

Teachers talked about student motivation as a major purpose of assessment; this was seen as primarily extrinsic in nature. Students were seen as being motivated in two ways: a) competition and pressure and b) enjoyment and praise. Competition and pressure were seen as motivating students to work harder; these were often related to the quantitative measurement of learning used for reporting to parents and the board (discussed more thoroughly in Categories #2 and #3). Good grades and qualifications (e.g. NCEA credentials) were described as a reward for hard work and sustained effort, but bad grades were thought to cause some students to quit trying. Teachers indicated that 'fun' assessment tasks motivated students and that the use of praise reinforced this positive affective feeling.

Most teachers saw assessment as crucial to maintaining student extrinsic motivation. For example, Fred explained, "...you heard the comment of "if I don't get excellence on this test, my mum's going to dah, dah, dah," so good motivational tools, assessments" (F1:176). Here, reporting to parents, as described in Category #3 is seen as motivating students out of fear. However, in addition to outside parental pressure to perform, many teachers cited that students have an internal desire to perform and achieve. For example, Vince explained:

A student wants to pass. That's why they're at school, that's why they do anything academic. You've got to pass. If you're not, you're failing and you don't want to fail. So unless you can judge, um, an informal assessment with a mark, they're not really interested. Rewards system I guess. (V1:226)

Especially the high school teachers noted that students became disinterested in class activities if marks or scores were not being awarded; while Vince talked about a student's internal desire to achieve, he discussed how this is based on an extrinsic 'rewards system.'

While commonplace in secondary schools, primary and intermediate teachers reported using few scored or marked assessments. Some within the sample felt that students really needed these as an incentive to achieve. For example, Alicia explained:

I have found that they need a sort of, a summative, a formal kind of test situation, contrary to what I have been told. For a while it was a swear word. We don't use tests and you're not allowed to. ... There's that part in every person that wants to feel that they've achieved something. That they've become, they've mastered it. (A1:044)

However, teachers like Tom disagreed that giving kids marks or scores positively motivated them:

... if you have straight assessment where peers, they all get a mark, then it does tend, it can be a negative influence on them... if it was a test and it was out of 10, then, you know, the ones who get 9 and 10 are the ones who are yahooping it and the ones who only get 4 or 5 are the ones who tend to think, "Oh, I'm not as good as they are." (H1:100, 102)

Grades or scores were seen as something positive for high achieving students, but negative for lower achieving ones. At the high school level, external qualifications

like the NCEA were seen as motivating students to achieve because they wanted to get this qualification.

In addition to competition and pressure, teachers said that students were also motivated when assessments were enjoyable and when teachers offered praise throughout the task. For example, Rebecca explained:

And so kids are really motivated to do things... they look at the fun side of it, but they don't realise they're actually putting together of their skills and they're actually applying themselves and there's got to be lots of team work and working together and we learn a lot from each other. (R1:026)

Assessments with a 'fun side' were as seen as motivating even when grades and marks are not given, with students often unaware they are even being assessed.

Teachers noted that students are also motivated when they are being praised. As Lisa explained:

...praise is an excellent way to keep them going. So if I walk past and I say "oh my gosh, look at that wonderful sentence" or "I love your impact beginning," or something then they're like, "Oh, I've got to write good," and they keep writing. (L1:100)

Praise and enjoyment were not seen as having a potentially negative side like competition and pressure.

Teachers said that extrinsically motivating students to learn and achieve was a major purpose of assessment. Mechanisms discussed in Categories #2 and #3 (Reporting to the board and to parents) contributed to student competition and pressure, while praise and enjoyment were viewed as more positive motivational tools. Grades and external credentials were considered to be rewards for those who worked hard and achieved well; the potential for comparison and perceived pressure implicit within these mechanisms were seen as making students work harder either because of fear of failure or the positive feeling of success. Despite concerns about the negative impact grades and marks can have on low achieving students, teachers were more positive about assessment being used to motivate students than purposes described in previous categories.

Category #5- Facilitating group instruction

Another teacher-reported purpose of assessment was to help schools organise teaching and learning effectively by facilitating group instruction; this and the final two categories (#6 and #7) were seen as leading to student improvement. For many schools, student ability grouping, either through streamed classes or differentiated small group instruction, was a part of their instructional practice; a combination of standardised test data and teacher input was reportedly used to make these decisions. Teachers also said they needed to prepare pupils for types of assessment they would encounter in future schooling and assessment their understanding of procedures designed to facilitate the organisation of their learning.

Within schools, teachers reported that assessment was frequently conducted to group students. In primary and intermediate schools, diverse grouping practices were built around assessment data including within and between class groupings and special pull-out extension and remedial programs. Alicia spoke of the important role

assessment plays in grouping students saying, “I’m not going to choose it [groupings] based on the colour of each child’s hair, so I have to do assessment. So it’s based for [the] practical implication of making a classroom program effective and manageable” (A1:068).

In primary school, teachers reported within-class groupings as fluid and, as the year progressed, based more on teacher observation than test results. However, even at primary and intermediate levels, participation in pull-out programs was said to be quite stable and based heavily on test data. These between class groupings were seen as important for the children, but also for the school’s image. For example, Rebecca explained:

... As a school you have to have a literacy support group so it looks good; we cater for all our kids. You’ve got to have this group and that group and we’re so busy trying to get the groups that I sometimes wonder whether we forget that we have to get those kids out of those groups as well. (R1:032)

While no teachers argued that grouping was negative for students, several noted that children’s academic needs changed throughout the year and that extension and support classes should be altered more frequently in response to individual children’s growth and decline.

At high schools, diagnostic standardised testing was frequently used to form streamed classes. While these placements rely primarily on the test data, teacher input was included as well. As Vince noted:

When it comes to placing kids, initially it’s taken on their test results. Then, if there’s any students that teachers think are wrongly placed, regardless of their test results, then teacher recommendation comes in as well. (V1:086)

Unlike assessment for reporting purposes (Categories #2 and #3), here teacher judgments were blended with test data to make decisions about student ability.

Teachers also reported using assessment to prepare students for practices they would likely encounter in high school and at university; all students were seen as needing this preparation because of current school structures. Primary and intermediate teachers that used tests frequently cited this as a purpose. For example, Chelsea, an intermediate teacher explained:

...we’re trying to prepare children ... it’s all very well having this formative assessment and “gosh you’re doing well here” ... but we’ve got to be realistic that there’s an exam world out there and we’ve got to prepare them for exams and if they go to Grammar, day 1 they have exams, day 1... We’ve got to prepare them for that. And also university’s all about exams. (C1:072)

While many teachers expressed dislike for some of these practices, they also said that if they didn’t begin using them with the students, their pupils would be disadvantaged in further schooling. Teachers of high school year levels 9 and 10 also reported using versions of senior assessment criteria with their students to prepare them for what was to come.

Teachers also reported a range of assessments to strengthen classroom routines, rules, and behavioural expectations in order to facilitate group instruction. For example, some teachers, like Emma, cited using assessment to check if students had done assigned work:

I call them quick fire quizzes... it's 5 random questions about the book... They should get all five, but if they can't even get one, it shows me that they haven't read the novel, although they are probably claiming to have read it. (E1:036)

This kind of assessment was designed to help teachers know if students have completed work and to encourage students to develop independent work habits. Teachers also talked about assessing the presentation of work and students' levels of preparation. For example, Oliver mentioned:

.... assessment is not just academic too. It's, you know, kids are still, even at year 5, missing 4 or 5 pages in a book and then starting the next page or they still aren't ruling up or they can't find a pencil after 16 weeks, you know. (O1:026)

These kinds of behaviours were considered important norms and expectations within the school that teachers were encouraged to monitor as they were necessary when conducting group instruction.

Teachers frequently reported using assessment to facilitate group instruction by grouping students within and between classes, preparing them for future assessments, and checking to make sure students followed protocols relating to academic work. Unlike Categories #1 and #2, a wider range of assessment types were used for these purposes, even though the practice of grouping students relied heavily on testing. Teachers expressed that they were trying to help students when introducing them to future assessment protocols and encouraging them to complete work and present it in standardised ways. However, throughout this category, the needs of students as a group were the focus, not the needs of individual pupils.

Category #6- For teacher use to personalise teaching and learning

In the final two categories, teachers began to talk about using assessment to personal learning in ways that would maximise improvement. While in Category #7, teachers described themselves as working collaboratively with students to diagnose individual pupil learning needs and act on these data, in this category teachers described themselves as solely responsible for these interpretations and actions. Teachers talked about using assessment data to plan future teaching, modify activities and assessments for struggling students, and to know when information should be retaught. While teachers in this category reported that they shared most assessment data with students, they did not expect pupils to interpret these data themselves; teachers said they corrected errors and gave feedback which students could use to improve. In this way, assessment was a primarily teacher-controlled exercise.

Teachers talked about using student assessment results to inform their planning and teaching; this was done to improve student learning. Assessment was seen as a teacher-controlled process. As Rebecca explained:

... I think it's [assessment's] more for us [the teachers]; the assessment is more for us, to help us work ways that are going to motivate and pick up the student's learning. It's not something where we give them a mark, and say, "Well, there's your mark; figure it out." It's us looking at the mark and thinking, "This child is way below; there's things we need to do to boost this child." And so I think that's the way we interpret it. Yeah, I think assessment is for us, it's a teacher's tool and the way we choose to use it will impact on

our student's development. I think it's for us, and I don't believe it's a student tool. (R1:066)

This reflected the more teacher-centred viewpoints articulated in this category; the high usage of singular personal pronouns like 'I' showed teacher ownership of the assessment process. However, unlike the previous category, singular nouns and pronouns are used when talking about pupils showing that the focus was on the individual children, not the collective group like in the previous category.

Teachers in this category reported using data to plan, modify instruction, and know when to reteach. For example, Nicole stated:

... if I've observed a particular student is struggling with the vocabulary of a unit, then um, for the next task or activity, I might give them less words to look up or provide more meanings or, you know, when you do the mix and match, I might leave some of theirs already matched up for them. (N1:050)

Teacher commonly reported using pre-tests to make sure work was at an appropriate level and post-tests to determine if learning had occurred; if end scores were too low some reported trying to reteach the unit or integrate the missing skills to the students who required them in future units. Assessment was also seen as providing teachers with feedback on the effectiveness of their teaching which they could use to change their teaching practices.

As assessment was viewed as primarily for teacher purposes, teachers indicated that they did not share all results with their students. For example, Henry stated, "I don't show [assessment results to] every child. They don't all see the results. They're there, but I usually take [them] when at parent interviews" (H1:044). However, most teachers within this category said they shared assessment results with students. For example, Wynona explained:

I try to make a point of once they've sat their tests and I've marked it and I give it back to them, we go over it. Some of them I do one to one conferencing, um just to say, "Well listen, you just missed out because of this, and this is why. Why did you do that?" So I find out what they're thinking and then hopefully I can steer them to the right track... (W1:030)

In this scenario, while the conferencing is done to help students understand mistakes, it is also clearly for her purposes, so she can understand the thinking behind individual students' mistakes.

In this category, teachers said the purpose of assessment was to help them individualise teaching and learning so students could improve; this was described in a teacher-directed way. Assessment was used to plan instruction, modify for individual student needs, determine when re-teaching is necessary, and evaluate past instruction. While teachers generally shared assessment results with students, this did not always occur because students weren't expected to make sense of their own assessment results.

Category #7 Teachers and students use assessment jointly to personalise teaching and learning

In the final category, while teachers also described using assessment to personalise teaching and learning, this was done through joint teacher-student interactions rather than the teacher-controlled process described in Category #6. In this category,

teachers reported working jointly with students to decide what the assessment data meant and how to best act on it; both share ownership of the data and its interpretation through a range of primarily formative practices.

Teachers here reported that assessment was for joint teacher and student usage. For example, Madison stated:

I'd like to think of it [assessment] as more of a joint thing... It has to be seen by the children and by me as a joint effort. We're going to find out how much we've actually learned, remembered, retained um, learned to think... There's not too much reflection on how hopeless we were or not, but where do we have to go next and I think that's the important thing. (M1:060)

This statement summed up the way teachers within this category described the purpose of assessment. The plural pronouns like 'we' are used rather than 'I.' Conferencing was said to be one of the best ways to assess as it allowed teachers and students to dialogue together about learning goals.

Teachers describing using assessment this way said it helped students become independent learners. For example, Grace stated:

... it's all about having a shared understanding, that shared vocabulary and going over it again, and again, and again until they do get it ... once they understand that and they start applying it, then they've got that, "Oh, I can see where I have to go next. I don't have to um, have the teacher tell me." And what we've just started doing now is like, "You tell me what you want me to do with the lesson? What are your weak areas that you don't understand?" so they've just started doing that. (G1:050)

Here Grace identified that she ideally wanted students to have enough understanding of their learning that they no longer required teacher direction. Teachers in this category reported encouraging their students to evaluate and judge their own work.

Here, teachers reported being highly accountable to their students. For example, Xavier explained:

... they [students] need to be interested in what they're learning about which means they need to be interested in how they're being assessed for it... they [students] do look at the surface features more, where teachers will look at the deeper features, but that's okay because if that's important to them, then it's important to us in the end because they're the stakeholders. (X1:072)

Teachers reported changing or modifying assessment criteria to make it aligned with what students considered important. When assessing student work, teachers often reported co-marking it with the child. For example, Madison explained, "... I mark it [their work] with them. They bring it to me and we talk about where to next..." (M1:022). Marking together in this way was said to be powerful as it helped students identify their own mistakes. Teachers also described student-led parent conferencing where pupils had an active role in reporting their achievement data to their parents.

However, some teachers in this category mention that their school culture or policies prevented them from fully including students in the assessment process. For example, Xavier explained:

We use rubrics here a lot. I don't think I use them well enough yet, because I think that the kids really need to have a lot of buy-in, but we sort of sit around a staff meeting and we sort of make it up ourselves, when really the kids

should have buy-in. “What do you think someone that learns very well does when we’re studying this?” And so they should be able to say, “They do this; they do that”. (X1:054)

In addition, teachers perceived that parents wanted teacher evaluations of their student’s learning, not opinions of the pupil.

In this final category, teachers described assessment as a process where they worked together with pupils to personalise learning in order to maximise student improvement. Students were said to have an active role when doing things like directing lessons, assessing their own work, assessing jointly with the teacher, and reporting assessment data to their parents. Assessment conducted in this way was seen as giving students ownership.

Discussion

Within this data set, teachers described seven major purposes for assessment. All teachers said some assessments were completed to comply with requirements (Category #1); teachers mainly discussed standardised testing within this category, but also included practices that went against their personal beliefs. These assessments were seen as inaccurate, unreliable, or negative for students. The next two categories (#2- Reporting to the board and external organisations, #3 Reporting to parents) were viewed more positively, but teachers articulated concerns about the quantification of learning for reporting purposes. A relatively narrow range of assessment types were described as used for reporting purposes including heavy use of testing. Categories #2 and #3 were viewed as related to student extrinsic motivation (Category #4) as the quantifiable student assessment data used for reporting could lead to competition between students and pressure to perform. While pressure and competition were seen as extrinsically motivating students, praise and enjoyment were viewed as more positive extrinsic motivators.

The final three categories related to assessment for improvement. In Category #5, teachers described using assessment to facilitate group instruction through with-in and between class grouping, preparing students for future assessment, and reinforcing procedural norms.

The final two categories focused on individualising teaching and learning, rather than catering to the needs of the whole group. While all teachers commented on the first five purposes (although not necessarily agreeing each was a valid and useful purpose of assessment), teachers split in the final two categories; While some reported that assessment was a tool for teachers (Category #6), others suggested that assessment must be used jointly by teachers and students for maximum improvement to occur (Category #7). In both Categories #6 and #7, teachers discussed using a wide range of formal and informal types of assessment. It is important to note here that the outcome space for teachers’ conceptions of assessment is much more complex than a simple dichotomy between formative and summative; an argument that Brown makes (2004, 2006b, 2008). Teachers’ conceptions of assessment are affected by multiple pressures. These include pressures like not doing harm to students, fulfilling obligations to employer, seeking to improve schools, and helping students to learn. These multiple tensions create a much more complex teaching thinking space than the simple dichotomy.

Within these data, three of the four main purposes for assessment clearly are evident in the data: improvement, school accountability, and irrelevance. Accountability was primarily defined as school accountability, not student accountability, although elements of student accountability are present in comments to do with preparing students for high-stakes qualifications assessments later in their schooling. While teachers also discussed grouping students (another aspect Brown [2002, 2004, 2006a, 2007] included in his 'student accountability' construct), they talked about it as being for learning purposes, not to keep pupils accountable. Many teachers within the sample rejected the notion of student accountability as it put undue pressure on children and could, if they did not perform well, result in negative affective consequences. While this is in contrast to Brown's (2006b, 2007) research with New Zealand teachers who indicated strong agreement that assessment should keep students accountable, it is consistent with his finding that this notion was correlated with irrelevance. Hence it may be that this sample of teachers has provided greater insights into that relational structure.

While there are obvious tensions between these assessment purposes, teachers also reported conflicts between what they perceived was good for their students and what they thought was best for schools. Teachers had difficulty reconciling these data as some practices, like standardised testing, were seen as potentially useful for schools as they could be aggregated to provide a holistic picture of achievement, but as negative and not necessarily accurate when examining individual student scores. It is worth noting here that in Brown's survey research he reported that school accountability assessment was positively correlated with improvement; it seems the interview comments again shed light on that correlation.

To illustrate the tensions teachers reported between what was best for students and schools, Figure 2 plots major constructs teachers discussed within the categories on two axes. The x axis represents a continuum of positive and negative value for schools, while the y axis represents positive and negative value for students. The field of the figure is further divided by two arcs which separate the three purpose-oriented conceptions of assessment. The bottom arc represents the irrelevance, the top arc is improvement and the middle space is accountability.

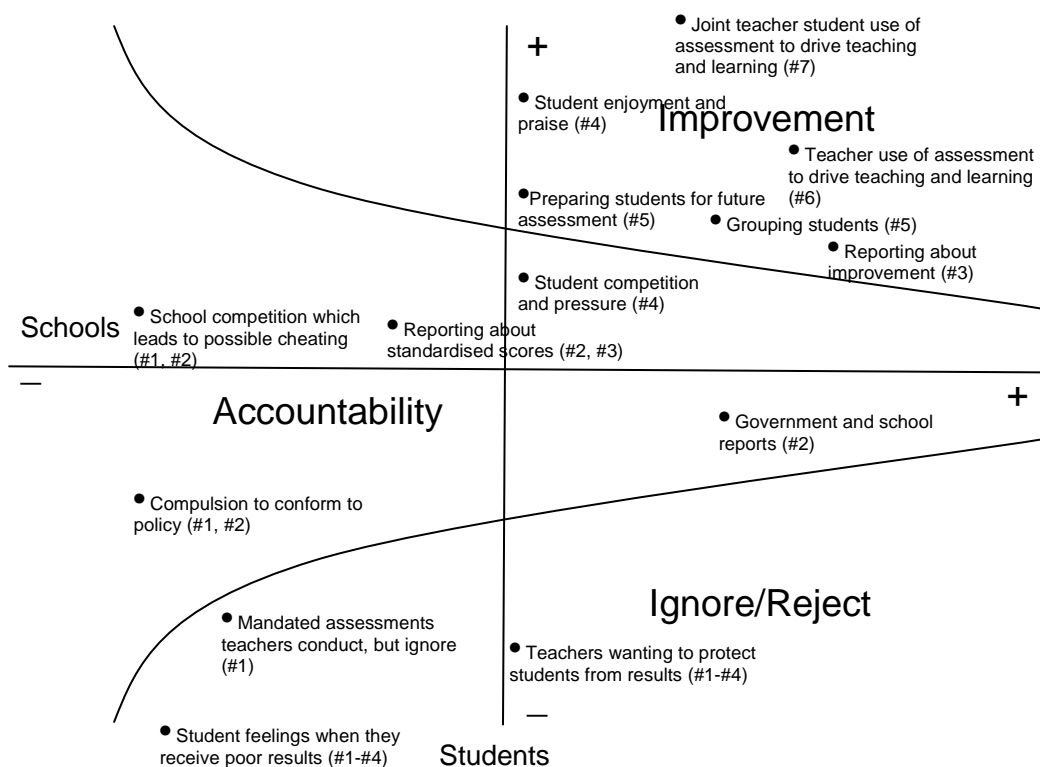


Figure 2- Relationships between the constructs within teachers' conceptions of assessment

Assessments that teachers reported ignoring or rejecting were seen as being bad for students and potentially negative for schools. The middle region captures the accountability-oriented categories around which teachers were slightly positive or negative. Finally, the improvement field captures categories that were deemed as positive for students and for schools as well. Most tension occurred within the accountability category, with teachers citing that practices like reporting to the board were positive and necessary for schools, but sometimes had negative consequences for students. Also, the relationship between Categories #6 and #7 highlights another tension. While including students in the process of determining and interpreting assessment was seen as extremely positive for student (Category #7), teachers recognised that students would not necessarily direct curriculum and assessment in the ways schools required. Category #6 was less positive for students than Category #7, but had fewer tensions with what was beneficial for the school as teachers controlled the agenda and could therefore accommodate curriculum and policy more easily.

These data confirm that teachers use a variety of assessment types for a range of purposes, some of which are more aligned with their personal opinions about assessment best practice. Teachers indicated that their students were their first priority and this led them to reject or ignore mandated assessment practices that they deemed as inaccurate (based on their own personal knowledge of the student's abilities) or negative for students. This is highly consistent with the child-centred pedagogy imbued in teacher practice in New Zealand (Fraser, 2001; Vaughan & Weeds, 1994)

While teachers who reported assessment as being for joint student and teacher use could be considered most aligned with the assessment *for* learning construct (i.e. Black & Williams, 1998; Clarke, 2005), these data highlight the problems teachers in current school environments experience when working in this way. For example, Vince noted the difficulties in enacting student centred practices in schools, saying, “You do student centred learning for a whole year and you might cover 1/5 of what you are supposed to” (V1:170). These data highlight the tension between ‘covering’ material (as required by the school and the ministry) and taking the time so students can own and direct the process.

Additionally, more empirical evidence is needed showing that the joint student-teacher practices described in Category #7 actually lead to more student improvement than teacher directed practices described in Category #6. For example, New Zealand secondary students have reported that self and peer assessment practices, often seen as central to the assessment *for* learning approach, are of little use as only teacher feedback is considered accurate and useful for improvement purposes (Peterson & Irving, 2008). There is a need for more research into how peer and self assessment practices could and should be effectively implemented in schools so all stakeholders view them as an important mechanism for improvement.

Notwithstanding the data coming from a small sample (26) of teachers drawn from Auckland, the pattern of conceptions revealed is consistent with previous studies with teachers of similar age ranged students. It would be interesting to investigate whether teachers of even younger or much older students have similar views. While the data are dependent on verbal self-report and analyst interpretation, there is a need for corroboration of the teacher views in practice. Of considerable interest would be the views of the teachers who refused the invitation to participate. Analysis of the TCoA-III questionnaire responses between the group willing to be interviewed and those refusing to be interviewed suggested substantial differences in their conceptions of assessment. Interviews would allow exploration of the reasons for their differing conceptions.

Another difficult issue is that teachers who reported the same conception of assessment did not cite analysing or using assessment data in the same way or to the same degree. For example, within Category #6, there is certainly a difference between teachers like Emma who said things like, “I suppose I use it for my lesson planning, and what’s going to be too difficult for these kids and all the rest of it” (E1:084) and teachers like Rebecca who described very systematic and thorough individualised planning based on their assessment data. Furthermore, it would be interesting to investigate whether the differences between Categories #6 (teacher-centred) and #7 (student-engaged) have differential impacts on student learning outcomes.

This study has shown that teachers see assessment as having a range of diverse purposes; different assessment practices are seen as being aligned with these. While assessment clearly has a range of purposes, it is important that teachers are enabled to move beyond compliance so that they can actually make educational use of data they collect, which is the only truly valid role of assessment (Popham, 2000). Clearly, misunderstanding of standardised tests and non-flexible school policies conspire against teachers exercising professional responsibilities around assessment.

References

- Airasian, P. W. (1997). *Classroom assessment*. (3rd ed.). New York: McGraw-Hill.
- Åkerlind, G. (2002). *Principles and practice in phenomenographic research*. Paper presented at the Current Issues in Phenomenography, Canberra.
- Ashworth, P., & Lucas, U. (1998). What is the 'world' of phenomenography? *Scandinavian Journal of Educational Research*, 42(4), 415-431.
- Ashworth, P., & Lucas, U. (2000). Achieving empathy and engagement: A practical approach to the design, conduct and reporting of phenomenographic research. *Studies in Higher Education*, 25(3), 295-308.
- 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. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Boud, D. (1995). Assessment and learning: Contradictory or complementary? In P. Knight (Ed.), *Assessment for learning in Higher Education* (pp. 35-48). London: Kogan Page.
- Butterfield, S., Williams, A., & Marr, A. (1999). Talking about assessment: mentor-student dialogues about pupil assessment in initial teacher training. *Assessment in Education*, 6(2), 225-246.
- Brown, G. T. L. (2002). *Teachers' conceptions of assessment*. Unpublished Doctoral Dissertation, University of Auckland, Auckland, NZ.
- 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. (2006a). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99, 166-170.
- Brown, G. T. L. (2006b). Integrating teachers' conceptions: Assessment, teaching, learning, curriculum, and efficacy. In A. P. Prescott (Ed.), *The Concept of Self in Education, Family and Sports* (pp. 1-48). New York: Nova Science.
- 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., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (in press). Assessment for improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy and Practice*.
- 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.
- Clark, C., & Peterson, P. (1986). Teachers' thought processes. In M. Wittrock (Ed.), *Handbook of research on teaching*. (3rd ed., pp. 255-296). New York: MacMillan.
- Clarke, S. (2005). *Formative assessment in the secondary classroom*. London: Hodder Murray.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.

- Dixon, H. (1999). *The Effect of Policy on Practice: An Analysis of Teachers' Perceptions of School Based Assessment Practice*. Unpublished Masters thesis, Massey University, Albany, NZ.
- Entwistle, N. (2000). *Promoting deep learning through teaching and assessment: Conceptual frameworks and educational contexts*. Paper presented at the TLRP Conference, Leicester, UK.
- 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.
- Fraser, D. (2001). Developing classroom culture: Setting the climate for learning. In C. McGee & D. Fraser (Eds.), *The Professional Practice of Teaching* (2nd ed., pp. 15-34). Palmerston North, NZ: Dunmore Press.
- Guthrie, J. T. (2002). Preparing students for high-stakes test taking in reading. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 370-391). Newark, DE: International Reading Association.
- Harris, L. R., Irving, S. E., & Petersen, E. (2008). *Secondary teachers' conceptions of the purpose of assessment and feedback*. Paper presented at the AARE International Research Conference Brisbane.
- Hamilton, L. (2003). Assessment as a policy tool. *Review of Research in Education*, 27, 25-68.
- Heaton, J. B. (1975). *Writing English language tests*. London: Longman.
- Hill, M. F. (2000). *Remapping the assessment landscape: primary teachers reconstructing assessment in self-managing schools*. Unpublished Ph.D. Thesis, University of Waikato, Hamilton, NZ.
- Jones, L. V. (2001). Assessing achievement versus high-stakes testing: A crucial contrast. *Educational Assessment*, 7(1), 21-28.
- Lankshear, C., & Knobel, M. (2004). *A handbook for teacher research*. Berkshire: Open University Press.
- Marton, F. (1981). Phenomenography- Describing conceptions of the world around us. *Instructional Science*, 10, 177-200.
- Marton, F. (1986). Phenomenography- A research approach to investigating different understandings of reality. *Journal of Thought*, 21(3), 28-49.
- Marton, F. (1994). Phenomenography. In T. Husen & T. N. Postlethwaite (Eds.), *The International Encyclopedia of Education* (Vol. 8, pp. 4424-4429). New York: Pergamon.
- Marton, F. (1996). Cognosco ergo sum- Reflections on reflections. In G. Dall'Alba & B. Hasselgren (Eds.), *Reflections on phenomenography: Toward a methodology* (pp. 163-187). Goteborg: Acta Universitatis Gothoburgensis.
- Marton, F., & Booth, S. (1997). *Learning and awareness*. Mahwah: Lawrence Erlbaum Associates.
- Marton, F., & Saljo, R. (1997). Approaches to learning. In F. Marton, D. Hounsell & N. Entwistle (Eds.), *The experience of learning* (pp. 39-58). Edinburgh: Scottish Academic Press.
- Mehrens, W. A., & Lehmann, I. J. (1984). *Measurement and evaluation in education and psychology* (3rd ed.). New York, NY: Holt, Rinehart and Winston.
- Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation*, 12(1), 53-74.
- Muijs, D., & Reynolds, D. (2005). *Effective teaching: Evidence and practice*. London: Sage Publications.

- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332.
- Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction*, 18, 238-250.
- Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders* (6th ed.). Boston: Allyn & Bacon
- Sandberg, J. (1997). Are phenomenographic results reliable? *Higher Education Research and Development*, 16(2), 203-212.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow, UK: Pearson Education.
- Sjostrom, B., & Dahlgren, L. O. (2002). Nursing theory and concept development or analysis: Applying phenomenography in nursing research. *Journal of Advanced Nursing*, 40(3), 339-345.
- Smith, M. L., Heinecke, W., & Noble, A. J. (1999). Assessment policy and political spectacle. *Teachers College Record*, 101(2), 157-191.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York: MacMillan.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Buckingham, UK: Open University Press.
- Vaughan, L., & Weeds, A. (1994). Managing an effective classroom. In C. McGee & D. Fraser (Eds.), *The Professional Practice of Teaching: An Introduction to Teaching, Learning and Curriculum* (pp. 131-174). Palmerston North, NZ: Dunmore Press.
- Walsh, E. (2000). Phenomenographic analysis of interview transcripts. In J. A. Bowden & E. Walsh (Eds.), *Phenomenography* (pp. 19-33). Melbourne: RMIT Publishing.
- Warren, E., & Nisbet, S. (1999). The relationship between the purported use of assessment techniques and beliefs about the uses of assessment. In J. M. Truran & K. M. Truran (Eds.), *22nd Annual Conference of the Mathematics Education and Research Group of Australasia* (Vol. 22, pp. 515-521). Adelaide, SA: MERGA.
- Webb, N. L. (1992). Assessment of students' knowledge of mathematics: Steps toward a theory. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 661-683). New York: Macmillan.
- Yu, C. (2003). *Meeting the workforce demands of Hong Kong's new era in secondary business education: Business teachers' conceptions of students' competence and choice of teaching strategies*. Unpublished PhD, Queensland University of Technology, Brisbane.