

ResearchSpace@Auckland

Version

This is the Accepted Manuscript version. This version is defined in the NISO recommended practice RP-8-2008 <http://www.niso.org/publications/rp/>

Suggested Reference

Brown, G. T. (2012). Teachers' thinking about assessment: Juggling improvement and accountability. *Teacher: the International Education Magazine*, 6(2), 30-35.

Copyright

Items in ResearchSpace are protected by copyright, with all rights reserved, unless otherwise indicated. Previously published items are made available in accordance with the copyright policy of the publisher.

<https://researchspace.auckland.ac.nz/docs/uoa-docs/rights.htm>

Teachers thinking about assessment: Juggling improvement and accountability

Gavin T L Brown, PhD

The University of Auckland

Assessment is everywhere in schooling. In some countries it begins before schooling even starts with screening tests and interviews for selection into nursery schools or kindergartens. For most children and adolescents, if it is not a regular occurrence, assessment looms in the background haunting or driving their learning. Assessment has many uses—selection, promotion, retention, deciding awards, grouping students, certification, reporting, tracking progress, and so on (Newton, 2007). Many of these uses are largely administrative rather than educational. While administrative demands are legitimate, in and of themselves, they are only indirectly related to the primary concerns of educators.

Educational uses and purposes focus much more on the idea that assessment can contribute to improving student learning and teachers' teaching. Popham (2000, p. 1) eloquently and forcefully stated:

if educational measurement doesn't lead to better education for students, then we shouldn't be doing it ... the only reason educators ought to assess students is in order to make more defensible educational decisions regarding those students. That's really why educators should be messing around with measurement-*to improve student learning*. [italics in original]

While Popham uses the term measurement, we can infer that he means assessment—a process of collecting information about student learning that leads to educational decision making. The key educational decision that every teacher needs to engage in is deciding *who needs to be taught what next* (Hattie & Timperley, 2007).

While educational improvement is an essential goal of assessment, this ambition is often overwhelmed by not just administrative, but also accountability uses. Accountability has to do with a simple idea: everyone has to give an account of what they have done in their work. In education, teachers are accountable for their students' learning and they are required to account for their effectiveness to their managers, leaders, and supervisors and, naturally, the parents of their students. A simple way, though not without problems, to evaluate teachers is to test their students—if teachers have done a good job, their students will do well on any assessment of what teachers were supposed to teach. However, there are many reasons students might not do well on an assessment, independent of how well the teacher has done his or her job. For example, the teacher may be working with students from poor homes, whose parents have little education and use a language different to that of the school. Alternatively, even an average teacher may appear very good, if the home resources of his or her students are high and those students are extensively helped out of school to learn for the assessment.

What makes accountability assessment important is that there are often negative consequences for schools and teachers who receive scores. In examination driven societies,

student assessment results are published in the media and used to determine the worth or quality of schools and teachers. In extreme cases, teachers have been fired, students have been forced to repeat grades, and schools have been disbanded. Being associated with high or low achieving schools brings reflected glory or shame to teachers. However, accountability that depends on tests or assessments is a two-edged sword. On the one hand, higher scores are meant to indicate that students have achieved what society expected from schools, and, on the other hand, high scores can be inflated without ensuring that the expected learning has taken place. Teaching to and cramming for the test, let alone cheating, can result in higher scores. While we might all wish to put our hands over our hearts and swear that no teacher would cheat, it seems highly probable that, if teachers conceive that the negative consequences associated with low scores are unfair or unethical, test score inflation practices will take place. These consequences are easily understood when we realise that accountability has fairly consistent and powerful effects on human—we tend to comply with the views of those to whom we must give an account of our work (Lerner & Tetlock, 1999). This means that we will produce results that meet what we know our superiors expect from us—and if the consequences of failure are extreme for the individual concerned (e.g., loss of reputation)—then we should expect teachers to deliver high test scores, through whatever means at their disposal.

When school systems place increasing responsibility for assessment in the hands of schools or teachers, rather than solely in the control of external examination boards, the thinking of teachers about the nature and purposes of assessment becomes much more important. If teachers conceive that assessment is primarily about evaluating student performance or if they conceive that learning from assessment (i.e., improvement) is largely what the student is supposed to do, then it is unlikely they will use assessment information to reconsider their own teaching practices, even if students in their class do poorly on one or more topics covered by the assessment. Offering teachers more training in test statistics, test item writing skills, or curriculum content knowledge, without addressing their conceptions of what assessment is and what is for, might simply create better skilled teachers who still think assessment is about and for the student rather than their own practice (Brown, 2008a).

Studies into teachers' thinking about the nature and purpose of assessment have been carried out in a number of countries using the Teachers' Conceptions of Assessment inventory (Brown, 2006; 2008b). This questionnaire asks teachers to indicate how much they agree or disagree with four competing purposes for assessment. The four purposes are: (1) assessment helps both teachers and students improve their teaching and learning respectively, (2) assessment evaluates or certifies student learning, (3) assessment evaluates or holds accountable schools and teachers, and (4) assessment is ignored since it has no meaningful purpose. For the sake of comparison, scores from a variety of studies have been put on a common scale where 1=strongly disagree, 2=mostly disagree, 3=slightly agree, 4=moderately agree, 5=mostly agree, and 6= strongly agree.

Table 1 shows the descriptive statistics for the four factors in the Teachers Conceptions of Assessment inventory from four different populations (i.e., Cyprus, Hong Kong, New Zealand, and Queensland). Officially, all four societies have low-consequence school

evaluation systems that do not rely primarily on formal testing to determine the worth or quality of schools. However, Hong Kong is a high-consequence, examination-oriented society, in which student evaluations formally determine life chances for children and, by inference, determine the quality of schools and teaching. Indeed, it seems that formative assessment in Hong Kong is treated as a soft policy, while examinations are the hard policy (Kennedy, Chan, & Fok, 2011). In the three other jurisdictions, student qualifications systems are run in secondary school, but three different approaches are taken. In Cyprus, student qualifications are determined by end-of-high school external examinations, in New Zealand, qualifications are determined by a mixture of in-school assessments and end-of-high school external examinations, and in Queensland, qualifications are determined only by teachers' in-school assessments. If teachers' thinking about assessment is shaped by how assessment is carried out in their society, then we should expect the scores for these scales to be consistent with the following hypotheses:

1. The Improvement purpose will be endorsed most strongly across all teacher groups. By inference, teachers will reject the conception of Irrelevance.
2. Student accountability will be more strongly endorsed than School Accountability.
3. Teacher conceptions will relate to the policy and practice priorities of their environment. For example,
 - a. Secondary school teachers will endorse Student Accountability more than primary teachers.
 - b. Hong Kong teachers will agree with School Accountability more than New Zealand or Queensland teachers.

Table 1. Teacher conceptions of assessment factor statistics

Country	Teaching Level	N	<u>Improvement</u> M (SD)	<u>Student</u> <u>Accountability</u> M (SD)	<u>School</u> <u>Accountability</u> M (SD)	<u>Irrelevance</u> M (SD)
<i>Brown (2011)</i>						
New Zealand	Primary	573	4.10 (.69)	3.54 (.95)	2.70 (1.01)	2.90 (.68)
	Secondary	404	4.02 (.66)	3.93 (.86)	2.68 (1.06)	2.90 (.65)
<i>Brown, Kennedy, Fok, Chan, & Yu (2009)</i>						
Hong Kong*	Primary & Secondary	288	4.76 (.38)	4.70 (.45)	4.09 (.62)	2.39 (.42)
<i>Brown, Lake, & Matters (2011)</i>						
Queensland	Primary	784	4.00 (.67)	3.64 (.76)	2.74 (1.12)	2.89 (.76)
	Secondary	614	3.84 (.66)	3.93 (.75)	2.69 (1.10)	2.87 (.69)
<i>Brown & Michaelides (2011)</i>						
Cyprus	Primary & Secondary	249	4.36 (.75)	—	3.82 (.81)	2.81 (.83)

Note. *Hong Kong data reverse scored and scale equated to the 6-point positively packed meanings.

Hypothesis 1. What is apparent from these surveys (Table 1) is the consistent finding that teachers endorsed the conception that assessment is for improved learning and teaching and

at the same time they consider assessment NOT to be irrelevant (low scores indicate disagreement). In other words, teachers universally believe in the notion that assessment can contribute to better teaching and learning. It may do so by motivating students to work harder to do well on a forthcoming assessment or it may do so by providing information to teachers about the learning needs of students. Clearly, from a formative perspective, assessments should help both the student learn better and the teacher teach better. What these responses suggest is that teachers believe assessment matters and that it is supposed to improve both teaching and learning. This is an important and powerful lever for introducing schooling improvement initiatives that make use of assessment. Teachers want students to learn more and they want to do a better job at teaching and, most importantly, they expect assessment to contribute to this. In other words, assessments that support diagnostic identification of student learning priorities will be highly valued by teachers. This creates both a challenge and an opportunity for assessment developers—how to put in the hands of those who are responsible for classroom learning tools that enable better teaching and learning.

Hypothesis 2. A further consistent pattern is that student accountability was endorsed more than school accountability in all groups where the two scales were available. Focusing on the learner as the purpose of the test is rational—we are testing the students, not the teachers, after all. While teachers may have qualms about evaluating students (i.e., mean scores were weak agreement), they seem to recognise that this function is a valid purpose. Where societies use assessments to identify merit and talent regardless of background or use qualifications systems as a transition mechanism into adult life, we should expect teachers, as members of that society, to endorse such processes. For some teachers student evaluation may be a mechanism for helping students develop and mature into self-regulating adults who take responsibility for their own learning and lives.

Hypothesis 3A. Another pattern was that primary and secondary teachers had similar views except around the use of assessment to evaluate students. The lower levels of agreement among primary school teachers in New Zealand and Queensland may reflect the policy and practice contexts of those two jurisdictions; in both places, there are no assessments that block student progression into secondary schooling. The very much higher level of endorsement in Hong Kong of this purpose is entirely consistent with the policies and practices of schooling in an examination-dominated society in which assessments are used from pre-school onwards to select and stream students into differentiated schooling opportunities.

Hypothesis 3B. Among all teachers in two jurisdictions (i.e., New Zealand and Queensland) the notion of using assessment to evaluate schools was fundamentally rejected; whereas in Hong Kong and Cyprus it was moderately endorsed. Furthermore, the correlation between school accountability and irrelevance tended to be negative, suggesting that this aspect of assessment is something they are likely to ignore. However, the survey alone does not explain why teachers are so cautious about school accountability as an assessment purpose. There are insights from teacher belief studies that throw light on this result.

- a) New Zealand primary teachers associated school accountability with deep learning assessments suggesting that assessments which do not show the deeper learning outcomes will be considered illegitimate measures of school effectiveness (Brown, 2009).
- b) When school assessment becomes predominantly focused on principal-led school-wide evaluation for improvement, teachers may see this practice as a form of external accountability and consider it irrelevant to their task of improving the learning of children in their own class (Brown & Harris, 2010).
- c) Among a large sample of teachers in Hong Kong and southern China, accountability purposes, which were correlated with irrelevance, incorporated the idea that examinations controlled teachers and schools (Brown, Hui, Yu, & Kennedy, 2011).

Thus, it seems that teaching professionals may consistently resent external surveillance and control as an inappropriate invasion of their educational prerogatives.

Another important insight into teacher rejection of external accountability examinations for teacher evaluation purposes has to do with educational reporting itself (Hattie & Brown, 2010). Most standardised assessments and examinations provide only total scores and/or rank order scores; that is, they tell teachers how much each person got right and their rank relative to other test-takers. While useful, this information is not sufficient to tell a teacher what the strengths of a low-ranking student are or what the needs of a high-ranking student are. The best students can still learn more and the weakest students are still competent to some degree at some aspects of learning. By constantly reinforcing rank order in test reporting teachers may develop a rather fixed-view of learning ability. Unfortunately, telling a teacher that a student is low-ranking may lead to the teacher thinking the student cannot learn more or better. Rather, we want teachers to have a flexible view of student learning capacity—all students can learn if taught...but it is difficult to believe this if assessments only provide comparison to the whole population rather than provide details as to what and how much each student has learned.

The problem for most teachers is that, while they may have been taught that they ought to analyse test performance for their own students, they do not have the skills or time to deconstruct test performance so as to guide changing their curriculum materials, teaching or learning activities, or grouping arrangements. Work in New Zealand has shown that rich diagnostic analysis of standardised tests (carried out by a computer-assisted reporting engine—Hattie & Brown, 2008; Hattie, Brown, & Keegan, 2003) has enabled teachers to efficiently identify the teaching and learning needs of their own students, respond appropriately, and see learning gains (Hattie & Brown, 2010). In Hong Kong, Carless (2011) has shown that teachers can be taught to use their own end-of-unit or end-of-term summative tests and examinations to formatively diagnose student learning needs and respond appropriately.

What these studies suggest is that teacher failure to use assessments formatively may reflect the deficiency of our own assessment systems, rather than an unwillingness to improve or change their teaching. It seems difficult to assume that teachers have failed to be formative

when assessment authorities and developers have not given teachers the tools they need to fulfil that purpose. Hence, there is a large resourcing change needed at the system level before teachers can be legitimately held accountable for raising student outcomes as measured by tests. Perhaps, teacher rejection of the school accountability purpose arises, not from a desire to inoculate their practices from evaluation, but rather from the lack of appropriate tests by which improvement and consequently accountability can be carried out.

Hence, if school-leaders and policy-makers are to introduce formative assessment or assessment for learning priorities, they ought to consider the pre-existing beliefs of teachers as a significant factor in determining whether their innovations will be implemented as intended. A number of important pre-conditions probably need to be met, if teachers are to be effective in using assessment formatively.

1. It seems important to give teachers the tools they need to accurately diagnose learning needs so that they can change their teaching practices. To do this would be to tap into teachers' predominant conception—assessment is for improved teaching and learning. Such tools, which may be most effective if drawing on information technology systems, need to give teachers quickly and effectively the information they need to know who needs to be taught what next.
2. The second condition, consistent with an approach that respects teachers' commitment to improvement, is to reduce the consequences of school accountability systems. This means, not removing such systems altogether, but rather adjusting their emphasis from a focus on amount achieved to a more subtle amount gained approach. Teachers may not be responsible for how much students achieve—there are many other factors at work—but they can be held professionally responsible for how much difference they contributed. This means not judging schools by how selective they were in admitting students, but rather in how much value they have added.
3. The third condition would be to allow sufficient time for real-world changes to be made—current research suggests that two to three years are needed for improved student learning outcomes to become visible. And the mechanisms used in New Zealand focus more on letting teachers see the problems themselves and helping them monitor the effect of their teaching interventions on student achievement. This approach taps into both teacher professionalism and teacher commitment to doing a good job.

The research reviewed in this brief paper suggests teacher thinking about assessment reflects the pressures and priorities of the system. Based on an understanding of teacher thinking about assessment, this paper has suggested approaches for designing assessments and school accountability systems. Since schooling improvement depends on teachers' professional capacity to monitor their own work, it seems logical to take into account their thinking about what assessment is and why it is implemented. Nonetheless, it should be kept in mind that teachers are generally committed to improved learning and teaching and need tools and conditions in which that commitment can flourish.

References

- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99(1), 166-170. doi: 10.2466/pr0.99.1.166-170
- Brown, G. T. L. (2008a). Assessment literacy training and teachers' conceptions of assessment. In C. Rubie-Davies & C. Rawlinson (Eds.), *Challenging Thinking about Teaching and Learning* (pp. 285-302). New York: Nova Science.
- Brown, G. T. L. (2008b). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science Publishers.
- Brown, G. T. L. (2009). Teachers' self-reported assessment practices and conceptions: Using structural equation modelling to examine measurement and structural models. In T. Teo & M. S. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (pp. 243-266). Rotterdam, NL: Sense Publishers.
- Brown, G. T. L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters*, 3, 45-70.
- Brown, G. T. L., Hui, S. K. F., Yu, W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, 50(5-6), 307-320. doi: 10.1016/j.ijer.2011.10.003
- Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy and Practice*, 16(3), 347-363. doi: 10.1080/09695940903319737
- Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, 27(1), 210-220. doi: 10.1016/j.tate.2010.08.003
- Brown, G. T. L., & Michaelides, M. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology of Education*, 26(3), 319-337. doi: 10.1007/s10212-010-0052-3.
- Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-Heritage settings*. London: Routledge.
- Hattie, J. A. C., & Brown, G. T. L. (2008). Technology for school-based assessment and assessment for learning: Development principles from New Zealand. *Journal of Educational Technology Systems*, 36(2), 189-201.
- Hattie, J. A., & Brown, G. T. L. (2010). Assessment and evaluation. In C. Rubie-Davies (Ed.) *Educational psychology: Concepts, research and challenges* (pp. 102-117). Abingdon, UK: Routledge.
- Hattie, J. A. C., Brown, G. T. L., & Keegan, P. J. (2003). A national teacher-managed, curriculum-based assessment system: Assessment Tools for Teaching & Learning asTTle). *International Journal of Learning*, 10, 771-778.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Kennedy, K. J., Chan, J. K. S., & Fok, P. K. (2011). Holding policy-makers to account: exploring 'soft' and 'hard' policy and the implications for curriculum reform. *London Review of Education*, 9(1), 41-54. doi:10.1080/14748460.2011.550433

- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255-275.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149-170.
- Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders* (6th ed.). Boston: Allyn & Bacon.