UNDERSTANDING TEACHERS' THINKING ABOUT ASSESSMENT: INSIGHTS FOR DEVELOPING BETTER EDUCATIONAL ASSESSMENTS.

Gavin T. L. Brown

Hong Kong Institute of Education

John A.C. Hattie

University of Auckland

Paper presented at the Annual Meeting of the National Council for Measurement in Education (NCME), San Diego, CA, April, 2009

Correspondence should be addressed to

Dr Gavin T L Brown

Dept of Curriculum & Instruction

Hong Kong Institute of Education

10 Lo Ping Road

Tai Po, NT

Hong Kong SAR

Email: gtlbrown@ied.edu.hk

UNDERSTANDING TEACHERS' THINKING ABOUT ASSESSMENT: INSIGHTS FOR DEVELOPING BETTER EDUCATIONAL ASSESSMENTS.

While teachers oftentimes design assessments for use in their classrooms, it is not the case that these teachers necessarily appreciate the psychometric qualities of such assessments or adhere to recommended standardised procedures for their use (Popham, 2000; Plake & Impara, 1997). For example, Impara, Divine, Bruce, Liverman, and Gay (1991) evaluated the ability of teachers to interpret a hypothetical student's test score information by providing reports that either had or did not have interpretive information. The availability of interpretive information and previous participation in a measurement class were significant predictors of the teachers' ability to accurately interpret score results. An overdependence on numerical information without interpretations was a major barrier; for example, most teachers were unable to correctly to interpret percentile band performance profile. The preference for interpretive information (i.e., analysis, synthesis, or the implications of test results) was also reported by Salvagno and Teglasi (1987).

Providing item information without appropriate interpretation can also be misleading and misunderstood. Linn and Dunbar (1992) showed that the anchor item exemplars of levels of achievement from the NAEP were easily misunderstood, and Hambleton and Slater's (1997) research on NAEP executive summary reports further highlighted the problems policy makers and educators have in understanding reports of assessment. Despite three-quarters of the sample had one or more statistics or testing courses in their background, "many interviewees had forgotten a lot of the statistical and measurement information they had known at one time" (Hambleton & Slater, 1997, p. 8). Problems were found with the statistical jargon or terminology used (e.g., "proficient", "statistically significant", "standard error", "cutpoint", "scale

score"), the construction of tables (e.g., cumulative columns exceeding 100%, too much detail, statistical symbols like greater than and less than signs, ordering of material, footnotes), and the design of graphs (e.g., over-complexity, novelty).

Nevertheless, as well as their self designed tests, teachers use a great many standardised, externally-developed, published tests as part of their assessment practices (Gipps, Brown, McCallum, & McAlister, 1995; McMillan, Myran, & Workman, 2002). In New Zealand, teachers reported using considerable numbers of standardised tests, such as the *Progressive Achievement Test* (Croft & Reid, 1991). They claimed that the school-based standardized achievement tests were used frequently or always to change how they taught their students (Croft, Strafford, & Mapa, 2000). Thus, as expected (Kane, 2006), teachers interpret the reports and/or scores from tests as part of their work and take educational actions based on their implicit and personal understandings of what the tests mean and are for. This proficiency to make use of tests (including design and selection) and other assessment methods to inform educational practice has been called assessment literacy (Quilter, 1998)

While it has been suggested that low levels of assessment literacy need to be remedied by better pre- and in-service professional development (e.g., Arter, 2001; National Research Council, 2001; Stiggins, 2001), a team of assessment developers in New Zealand has taken a more user-oriented approach (Hattie, Brown, Ward, Irving, & Keegan, 2006). Instead of requiring teachers to be more like test developers in terms of understanding the technical aspects underlying tests, the team has been investigating teachers' thinking about assessment to better develop test systems that communicate clearly and appropriately to teachers such that valid interpretations and actions are enabled. In this paper, we shall focus on results from surveys of New

Zealand teachers' conceptions of assessment, the implications for test development, and highlight some of the features of the test system that address those conceptions.

Teachers' Conceptions of Assessment

Teachers appear to have four major options as to the purpose of assessment (Brown, 2008; Heaton, 1975; Shohamy, 2001; Torrance & Pryor, 1998). The first conception is that assessment leads to educational improvements in terms of the teacher's teaching and the student's learning—this is frequently expressed as assessment *for* learning or formative assessment (Black & Wiliam, 1998). Even more radically, it can be expressed as assessment *as* learning. This is the dominant reason for schooling—helping students learn, understand, or do more than they could before and may be the only legitimate purpose of educational assessment (Popham, 2000). A second purpose is the use of assessment as a means of making students accountable for learning or certifying student achievement—this is frequently expressed as assessment *of* learning or summative assessment. This conception arises from the practice of evaluating student progress or achievement or ability and implementing consequences (e.g., changing track, awarding certificates) for the individual student as a result of assessment performance (Guthrie, 2002). A third purpose is the use of assessment as a means of holding schools and teachers accountable for achieving learning outcomes or standards—this is frequently associated with national testing or school accountability policies (Firestone, Mayrowetz, & Fairman, 1998). Such policies and practices are founded on the dual notions that teachers require external pressure to ensure students meet expected standards (Hershberg, 2002; Smith & Fey, 2000) and/or that teachers and students improve when compulsory external tests assess expected standards (Resnick & Resnick, 1989). The final purpose is that assessment is purposeless—this notion generally appears as a negative response to

both forms of accountability or the use of formal assessment tools for improvement. This irrelevance of assessment is founded on the notion that teachers do not need assessments to know what students need to be taught next, and they certainly do not need externally-imposed assessments.

A survey questionnaire (*Teachers' Conceptions of Assessment* (TCoA-IIIA) (Brown, 2002) has been used with primary and secondary teachers in New Zealand, Australia, and Hong Kong. The TCoA inventory generates four scores—one each for the four conceptions described above. In each population, an inter-correlated, four-factor model has been found to have adequate to good fit to the data (Brown, 2004, 2006, 2007; Brown, Kennedy, Fok, Chan, & Yu, in press; Brown & Lake, 2006), while differences in emphasis within each population have been detected. New Zealand and Queensland primary and secondary teachers (where there has *not* been a history of national mandated testing) generally agreed that assessment was about improvement and student accountability, while disagreeing with the irrelevance and school accountability conceptions (Brown, 2007; Brown & Lake, 2006). The secondary teachers agreed somewhat more with assessment as student accountability, but for the other three conceptions there were no differences.

For these teachers student accountability was positively correlated with irrelevance (r=.40, .21 primary and secondary respectively); whereas school accountability was associated with improvement (r=.41, .42 primary & secondary respectively). While teachers disagreed with the conception that assessment should make schools accountable, they associated their use of assessment with improvement of schooling. In contrast, the teachers agreed that assessment held students accountable but conceived such a conception as a bad thing. Hence, it would appear that, from the teacher perspective, a quality school generates improvement, but

schools should not be judged by assessments. Indeed, while teachers accepted that assessments are used to evaluate students, this is something that should not be done. Interestingly, Hong Kong teachers (Brown, et al., in press) had a very strong correlation (r=.91) between improvement and student accountability, suggesting strong cultural effects in how assessment is construed. Nonetheless, if assessments could provide information to schools as to how improvements could be made, rather than simply provide information about students, then it is likely the New Zealand teachers would be inclined to make use of them.

A sub-sample of about 240 primary teachers completed a frequency of assessment practices survey derived from McMillan (2001) and the TCoA. Four assessment practices factors were found related to format and cognitive demand; formal, test-like formats were contrasted with informal, interactive formats, while deep cognitive processing was contrasted with surface cognitive processing. Brown (2009) found that while the student accountability conception predicted the use of both surface cognitive processing (β=.33) and test-like formal (β=.45) formats, the school accountability conception predicted only the deep cognitive processing format like (β=.18). These structural relations were interpreted as evidence that teachers thought "*if you want to hold a school accountable, examine the impact teachers make on students' deep learning not the surface processing measured by externally-provided, student accountability tests"* (Brown, 2009, p. 263). Additionally, both the improvement and irrelevance conceptions of assessment predicted greater use of informal, interactive assessment formats (β=.57, .45 respectively), suggesting that formal tests are problematic sources of information.

A similar sized sub-sample completed the TCoA and selected self-report inventories on teaching perspectives, curriculum orientations, learning approaches,

and teacher-efficacy (Brown, 2008). A factor analysis of 21 scale scores found four integrated meta-factors. The two accountability conceptions of assessment were linked with transmission teaching, surface learning and internal efficacy in the summative accountability meta-factor, while the improvement conception was part of the formative assessment meta-factor along with apprenticeship and nurturing teaching perspectives, humanist curriculum orientations, and deep learning approaches. These two meta-factors were both viewed positively, with more agreement given to the formative improvement meta-factor. This analysis suggested that New Zealand teachers agreed with both of the major competing pressures in educational assessment—formative and summative—but these were associated with different teaching, learning, and curriculum conceptions. Accountability was seen as part of surface learning and teaching as telling; while improvement was seen as nurturing for children and deep learning.

Together these studies indicated that New Zealand teachers were mistrusting of assessments as means of demonstrating school quality, disliked evaluating students, associated ‘testing’ with making students accountable for surface learning, and believed that tests of deep learning could indicate school quality. From these studies we can also conclude that assessment systems need to report information about (1) deep aspects of learning, (2) means of improving teaching, and (3) ways of permitting schools to monitor their own effect. Further, systems should emphasise school- and teacher-based uses instead of externally, mandated accountability uses. It may also be concluded that if teachers appreciated the reporting value of the system, the formal nature of tests might not be an obstacle to their use.

The Assessment Tools for Teaching and Learning (asTTle) system

The New Zealand Ministry of Education commissioned the design and deployment of a standardised, computer-assisted testing system—Assessment Tools for Teaching and Learning (asTTle) (Hattie, Brown, & Keegan, 2003). The asTTle system, first released in 2002, and now in its seventh version, allows teachers to create their own standardised tests from banks of IRT-calibrated items in reading, mathematics, and writing in either English or Maori (Hattie, Brown, Keegan, et al. 2004). The system, instead of introducing new kinds of testing as suggested by the National Research Council (2001), introduced a new way of communicating assessment reports and channelling assessment information to teachers. Through an incremental research and development process (Hattie & Brown, 2008), adjustments were made with each version of the software to increase the range of options available to teachers and to ensure the effectiveness of the test reporting system. Further, the use of asTTle is not compulsory, nor are schools required to report results to the Ministry or other agencies; it is truly a school-based educational assessment system.

asTTle took up the challenge of teachers' thinking about assessment by developing reports that provided teachers both school accountability, student accountability, and improvement information through a series of interactive reports rather than trying to create one single report for all purposes. The major design requirement of Project asTTle reporting was to provide teachers with high-quality, externally referenced information about the achievement of students, as well as detailed, reliable information about the nature of student achievement. Thus, reporting outputs had to fulfil both normative and criterion referenced interpretations. Further, the outputs would initially be encountered in a graphic online environment that teachers could explore before determining which information would be printed or

transferred to school databases that are used for management purposes. This meant that an interface with some dynamic features and with high-quality user-interface characteristics (Spolsky, 2001) had to be developed.

Furthermore, the reporting system had to serve the purposes teachers had for assessment and do so in a way that added-value to the work teachers had to do. This motivating requirement (i.e., tell teachers something they did not already know; help them do their work more efficiently and effectively) was especially important because use of the system was voluntary. Additionally, the assessment system was developed as a curriculum-based educational improvement resource; use of the system was intended to improve the teaching and learning of mathematics, reading, and writing in both languages of instruction in New Zealand (i.e., English and Maori).

Consistent with recommendations from Wainer et al. (1999), pilot testing of visual displays was carried out with intended users. Focus groups were used to discuss issues pertaining to preferences for the design of the visual displays and the nature of intended users' understanding (Hattie, 2009). Procedures for testing the reporting design followed Scriven's (1991) procedure of internal alpha testing, including project personnel and Ministry of Education personnel, and external beta testing. Additional constraints on the design of the output existed in the form of programming requirements and budgets.

Reporting for School Accountability

For school accountability purposes, there are reports that allow teachers and school leaders the ability to compare student performance in similar types of schools ("School like Mine", Hattie, 2002). This report, called the Console Report (Figure 1), permits comparisons of the local school performance against national norms or against clusters of similar schools matched on location, rurality, size, socio-economic

status, and student ethnicity mix. The report permits fair comparison of like with like, while removing from teachers the potential excuse that poor performance is related to the student population characteristics. At the same time, the report allows schools to show that they are doing as well as or better than similar schools and hence demonstrate quality and accountability. In this way teachers can establish relativity of their group to appropriate contrasts—are my students above, at, or below other classes or schools who have students like mine? If the group is below these norms, the teacher is encouraged to take responsibility rather than blame students. If the group is above norms, the teacher is encouraged to extend student capability—something possible given the vertical indexing of performance to curriculum levels that extend from Level 2 to Level 6 (approximately 10 years of growth).

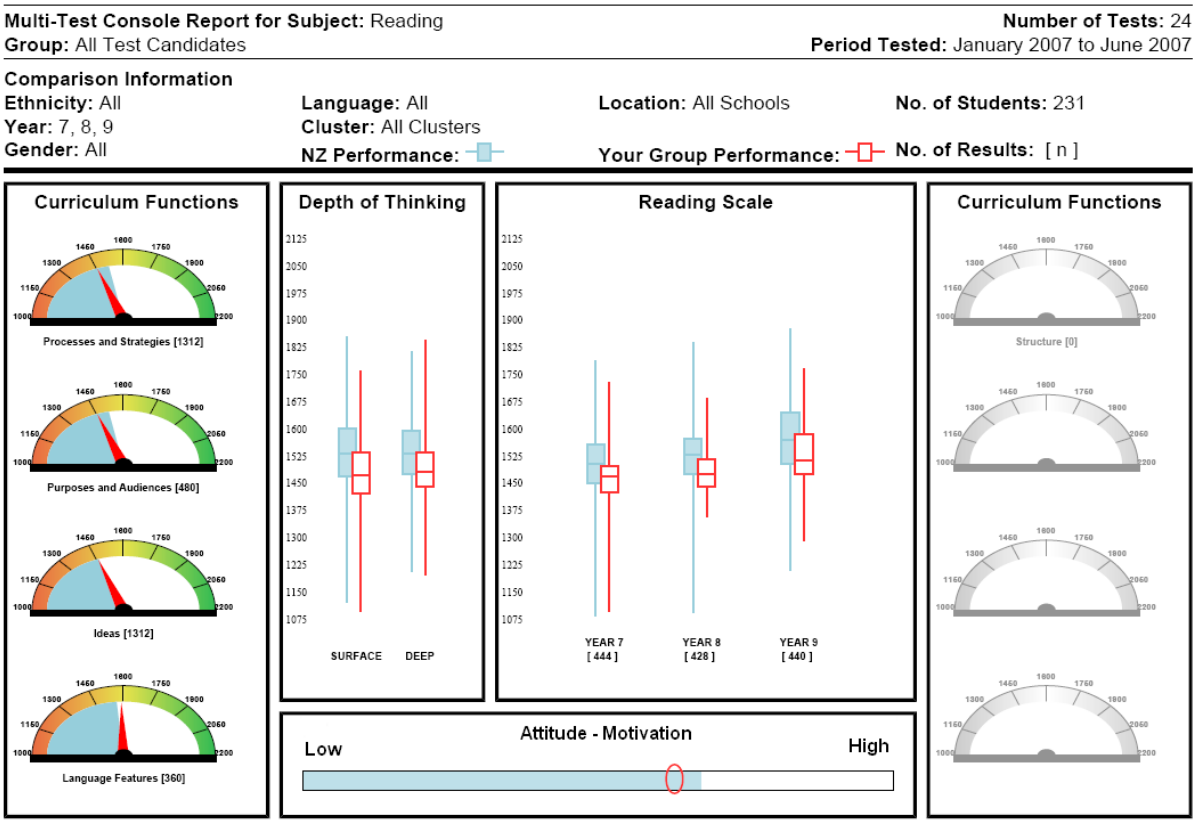

Figure 1. asTTle Console report

Another feature of the Console Report identifies how well students do on tasks that require surface or deep cognitive processing as defined by the Structure of

Observed Learning Outcomes (SOLO) taxonomy (Biggs & Collis, 1982). The SOLO taxonomy classifies tasks and responses that require single or sequential processing tasks as surface, while those that require integration of information or abstraction to more general principles as deep. Each test item was classified as surface or deep through content analysis by panels of teachers. Each test is required to include a minimum of 25% deep or surface items and performance within each type of cognitive processing is provided on the Console Report. This functionality provides information about how well students perform on cognitively demanding tasks relative to appropriate norms.

The Console Report also permits identification of year trends and performance across multiple tests—longitudinal growth and target-setting are features of e-asTTle currently being deployed full-scale in New Zealand (there is provision for a separate progress report over time for a student, class or school). Together these features of the Console Report allow schools to monitor their own effectiveness and demonstrate accountability against national norms and expectations. This takes advantage of the positive correlation teachers make between school accountability and improvement conceptions of assessment, as well as the association they make between deep learning and school accountability.

Reporting for Student Accountability

Student accountability purposes are met primarily through the Individual Learning Pathways report, a report which combines both ‘kid-map’ strength and weakness analyses with a personalised comparison to grade-appropriate norms. Teachers can use this report to share both with students and parents how well the child has performed and what aspects of learning students might need to pay attention to. Further, the report also provides an indication of what new learning the student will be

directed to. The Individual Learning Pathways Report permitted diagnostic description of student strengths and weaknesses at the curriculum objective level such that both teacher and student could identify what the next teaching or learning step was (Figure 2).

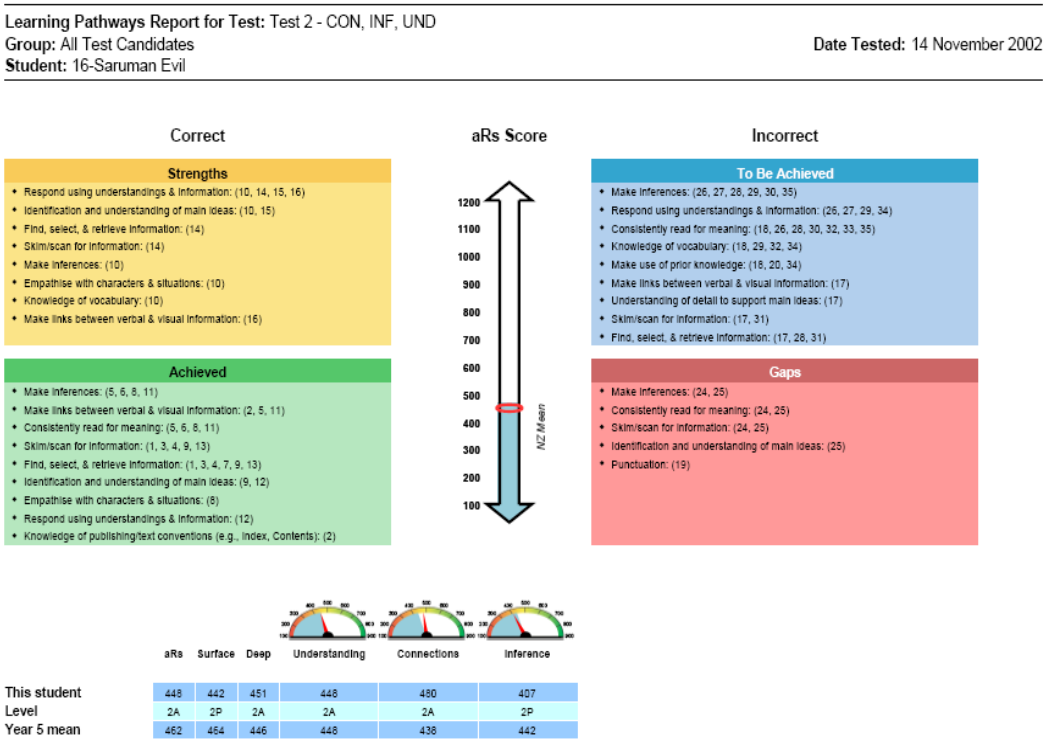

Figure 2. asTTle Individual Learning Pathways report

The ILP report is a variation of the IRT kid-map but instead of focusing on items, it converts the information into curriculum-defined achievement objectives that are the key interest and concern of teachers. Further, those curriculum-based objectives are the basis on which curriculum materials are developed, so this reporting encourages greater alignment between curriculum, assessment, and teaching. But the primary purpose of this information is to allow summation of the student's current status and identification of appropriate action on the part of the learner as well as the teacher. This report transforms simple rank-ordering test reporting into an evaluation of the individual that informs improved teaching; the essence of formative assessment to which New Zealand teachers are deeply committed.

Reporting for Improvement

Improvement purposes are met through group versions of the individual learning report, the curriculum levels report, the console report, the progress report, and the what next website. The Group Learning Pathways report identifies the areas of each curriculum that are major learning priorities for more students, even when they take different tests. The Group Learning Pathways Report (Figure 3) aggregated the diagnostic information such that priorities for teaching and learning could be determined across a group, cohort, or whole school population. While the ILP is a resource often used for parent-teacher conferences (and there are NZ schools using those reports for student-led conferences), the teacher and department head are responsible for setting pedagogical priorities for whole cohorts (e.g., class or year group).

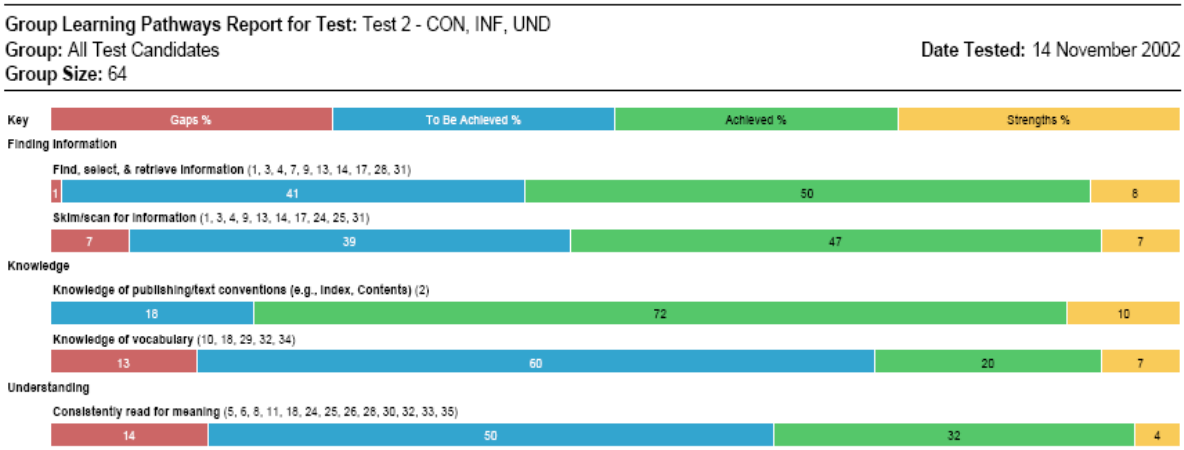

Figure 3. asTTle Group Learning Pathways report

The Curriculum Levels report identifies which students are at each sub-level (i.e., Basic, Proficient, and Advanced) for five of the eight national curriculum levels (typically covering Years 4 to 12). The Curriculum Levels Report (Figure 4) aggregated student performance by curriculum level sub-groupings determined by a reputable standard setting procedure (initially examinee- and item-centered methods were used but phased out in favor of the bookmark method) and which permitted

naming of students in each sub-level group so that appropriate educational materials and activities could be assigned to students at similar levels of progress. By selecting a chart, the teacher is provided a list of students with similar progress and learning needs. This permits within class or within cohort grouping. Given that performance in the multiple strands of the curriculum is not purely a function of general mental ability, students can be regrouped accurately depending on content.

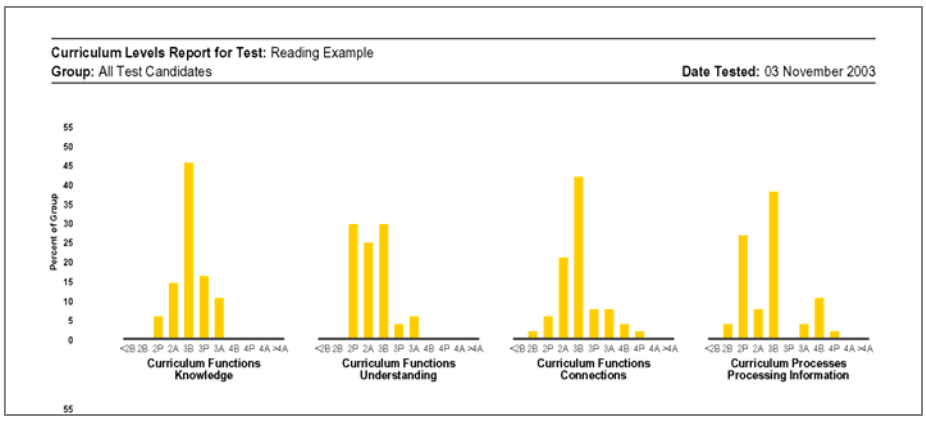

Figure 4. asTTle Curriculum Levels Report

The What Next website provides teachers (and students) a catalogued selection of high-quality teaching resources which can be used to address identified learning priorities. The What Next Report (Figure 5) provided a linkage from current performance levels to a web site that provided an indexed catalogue of teaching and classroom resources for each curriculum sub-level and each curriculum category used in the asTTle test creation system. This linkage further embeds the curriculum-focus of asTTle. This system is not a testing or computer system first-and-foremost; rather, it is an educational curriculum-resource that happens to use computers and testing. Linking to other curriculum (and government) resources gives greater power and value to a variety of development activities—a bonus when resources for improved teaching and learning are contested.

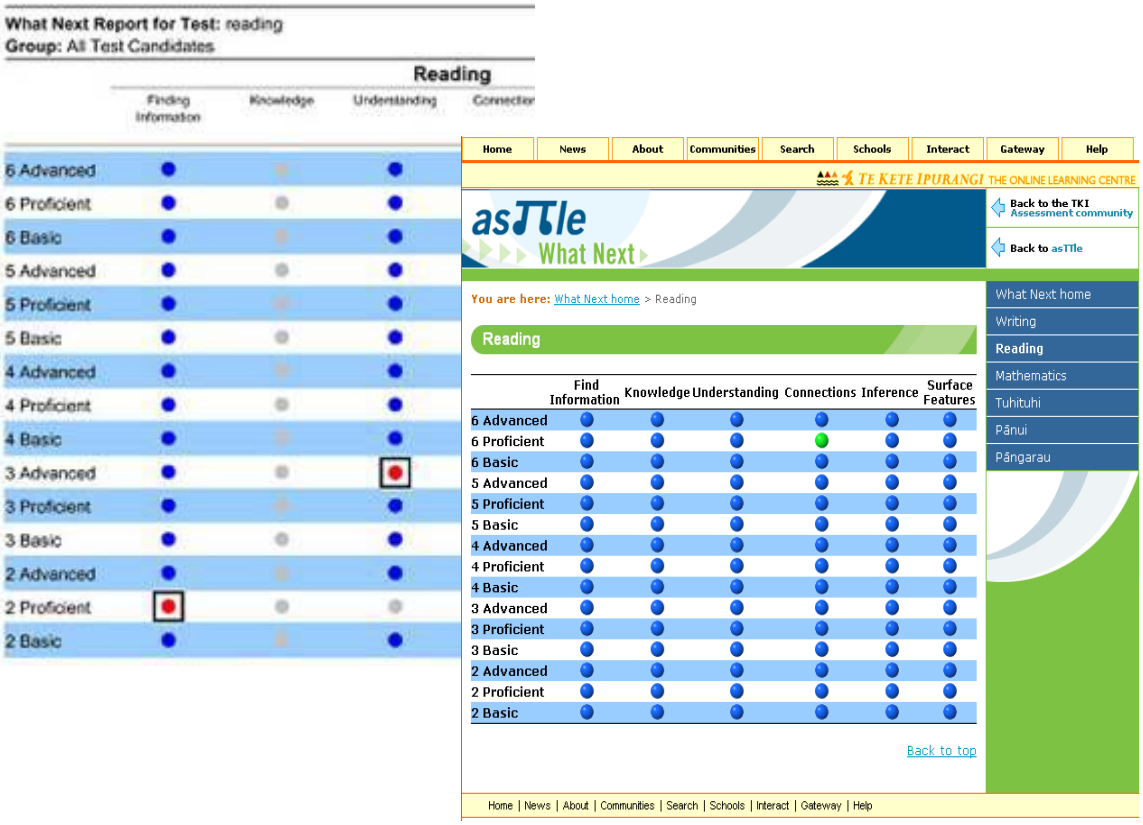

Figure 5. asTTle What Next report and asTTle What Next website

Together these reports address the conceptions teachers have of assessment. The system clearly provides for improvement-oriented analysis of student abilities (i.e., it identifies students' cognitive processing, guides teachers to materials for improved teaching, identifies who needs to be taught what). At the same time, it enables teachers to meet accountability requirements by enabling school leaders to obtain early warning indicators of how the school is doing and by helping teachers to explain clearly to parents and students how each child is doing relative to curriculum standards and norms. Within the context of low-stakes associated with using asTTle, the system permits teachers to evaluate accurately student learning and use that information to improve the quality of the school's effectiveness. Hence, it takes advantage of teachers' willingness to associate improvement with school accountability and with deep learning.

Effects of asTTle system on educational practice

Early evaluations established that teachers who received professional development in using the tool for educational improvement made more accurate interpretations of the reports (Hattie, et al., 2006). Upon release of asTTle Version 1 to a pilot sample of 110 New Zealand primary schools in mid-2002, a systematic evaluation of Year 5 to 7 teachers' understanding of the asTTle reports was conducted (Ward, Hattie, & Brown, 2003). To that end, a comprehension test, partially inspired by Hambleton and Slater (1997) and Linn and Dunbar (1992), of asTTle reports was created and administered as part of a survey questionnaire delivered to participants. A series of three questionnaires were presented in a matrix sampling pattern (each teacher received two of the three questionnaires) that asked about the interpretation of the asTTle reports. For example, a Console Report (based on simulated data) was presented and teachers asked "In which learning area did this class of students get the highest score?" or an Individual Learning Pathways Report was presented and teachers asked: What is the best way to understand the difference between Julie's asTTle Literacy Scale Score and the NZ reference group Year 5 mean?" Altogether there were 35 questions answered by 193 participants. The estimate of reliability of these items was .93, indicating that we can meaningfully interpret these items and the total score.

The overall mean for the Console Reports (.67 on a 0 = incorrect to 1 = correct scale) and What Next (.78) indicated a higher level of correct interpretations, but the means were lower for Individual Learning Pathways (ILP) (.51) and for Curriculum Levels (.57). Five of the items for ILPs were the lowest, particularly relating to writing. A major part of the misinterpretation was that too many teachers were interpreting the concepts and items in each of the cells (i.e., Gaps, Strengths, and

Achieved) relative to the NZ or class norms and not relative to the individual student's average ability. The concept that they should have understood is that relative to this particular student's ability, here are the concepts and items that he/she should have got correct (i.e., items relatively lower than this student's average proficiency in writing or reading) but did not (gaps), etc. Teachers, instead, were incorrectly interpreting the concepts in relation to the overall class or NZ norm group (these reports subsequently were modified in light of these findings). The major correlates (*r* ranged between .17 and .36) of interpreting the asTTle reports correctly were eight positive attitudes towards asTTle and its reports, including ease of use. Those teachers with a conception of assessment related to "assessment is powerful for improving teaching" had higher interpretation scores (r = .34), whereas those who had a conception of assessment as related to school accountability had the lowest interpretation scores (r = -.21). As argued by Brown (2004) professional development needs to attend to the conceptions of assessment held by teachers before introducing asTTle, because those who see assessment as being about school accountability rather than improvement of teaching are less likely to accurately interpret, or attend to the information in educational assessment reports.

Brown and Harris (2009) surveyed about 160 Auckland region teachers in 2008 concerning their conceptions of assessment as part of a study into teacher assessment practices. The teachers surveyed had a mean score for the conception that assessment was for school accountability that was considerably higher (d=1.86) than the two nationally representative samples of primary and secondary teachers surveyed in 2001 and 2007. The same group agreed more that assessment was irrelevant (d=.68), about the same that assessment was for student accountability (d=-.15), and much less that assessment was for improvement (d=-.83). On the whole, the

interviewed teachers were characterised by agreement that assessment was for school accountability. Thus, considerable shifts had taken place in the minds of teachers—perhaps this was a consequence of increasing pressure to report school effectiveness or else it could have been a function of increasing ability to monitor school effect through tools like asTTle.

Within the same study 26 teachers were interviewed, and it was concluded that the provision of asTTle and related professional development resources allowed teachers to equally address the formative improvement conception and use of assessment for accountability. For example, Alicia, a Year 8 (i.e., students approximately 12-13 years old) middle school teacher explained:

> the data [from a school wide test] is being used to report to the Board of Trustees. It's good for me because it's a formative assessment for me. I can use this to see how to, what I need to do to in order to set up my program for narrative writing, … that same information is used to report to the board.

She explained that other teachers in her school were taking data produced primarily for accountability reasons (i.e., reporting to the Board) and using it for their own formative purposes to improve teaching and learning within their own classes.

Similarly, Madison, a 6th grade (i.e., ~10-11 years old) primary school teacher, identified positive improvement-oriented benefits from asTTle:

> As a teaching tool, I think it [asTTle] is quite useful. But we discovered that by letting the children have just the limited 40 minutes that your slower children are not finishing and asTTle reports that what was not finished was not known, which is not necessarily true of course. So what we're trying to do is have that because that's a requirement for reporting, so we do that absolutely accurately, but then we give them another test which we allow them to go to the end of and then we use that for teaching because that will tell us what the children don't know

By modifying the formal accountability-oriented asTTle testing conditions, Madison was able to obtain formative information while meeting reporting requirements. Alicia

and Madison's examples illustrate that teachers are able to exploit the multiple features of asTTle to improve teaching and learning while still maintaining accountability.

Archer (2009) visited a NZ middle-school where students were aware of and appreciated the grouping and regrouping of students that the teacher implemented by content area. The students, had access to their asTTle reports. They believed they were learning and were not ashamed of being in a lower group in one content area as they knew they were getting appropriate materials and instruction depending on their observed performance on asTTle. The teachers indicated that they had adopted the asTTle system because of its multiple reports, which permitted access to improvement-related information rapidly, allowed appropriate differentiation of curriculum-aligned instruction, and allowed accurately monitoring of learning progress.

By seeing patterns across items and tests, it is possible to focus on curriculum objectives deemed important by the teachers and monitor the effectiveness of progress over a school-year. A New Zealand high school mathematics department has, through systematic school-wide monitoring of asTTle test results, been able to convert high school entry scores that were below the national average in Year 9 (Form 3) into above average national qualifications results in Year 11 (Form 5) (school principal, personal communication, June 2008).

Another program of research has focused on the professional development of teachers' literacy (i.e., reading and writing) instruction in 13 schools over a two-year period (Parr, Timperley, Reddish, Jesson, & Adams, 2007). The asTTle curriculum analysis was used as the framework for teachers' professional development in reading and writing instruction and the asTTle tests themselves were used as a means of

evaluating student learning outcomes. Students exhibited considerable gains in writing (*d*=1.28) and moderate gains in reading (*d*=.48). Thus, use of the asTTle system led to improved pedagogical content knowledge, especially in writing, which in turn led to greater student learning gains. Specifically, it was found that teachers improved in their ability in this project to use data from test reports to make data-referenced interpretations and recommendations (Parr & Timperley, 2008). Furthermore in an independent evaluation of the same project (McDowall, Cameron, Dingle, Gilmore, & MacGibbon, 2007), gains in student learning outcomes were associated with teachers who believed they had strong abilities to use and interpret assessment tools and greater knowledge of literacy. It was also shown that teacher development depended very much on the increased skills of the literacy professional developer herself. It would appear that the asTTle framework provided a coherent framework by which many schools could develop a common understanding of what literacy meant, what literacy pedagogy involved, and what improvement-oriented interpretation of student learning assessment results required.

There is an increasing body of evidence for the positive, intended effects of the asTTle testing system on teacher assessment practices. The reports align, we argue, with teachers' conception, and official government policy, that assessment exists to improve student learning and, further, that improvement is associated with demonstrating school accountability, rather than just measuring students' abilities. We believe that by enabling both improvement and accountability conceptions in its reporting system, the one system is able to support both governance and pedagogical goals in an educationally robust fashion. The studies of asTTle use have shown that how assessment is conceived and the beliefs that teachers have about assessment are associated with gains in student learning as well as more effective use of test reports.

Hence, we suggest that the New Zealand example demonstrates that if test development takes into account the pre-existing conceptions of teachers about assessment, it will result in test reporting and professional development that are more effective in raising student achievement. This is so because teachers will be able to use the tests for improvement, while satisfying accountability-oriented requirements. Taking into account both of these purposes for assessment and devising an integrated reporting system that addresses them appropriately is an essential aspect of assessment *for* and *of* learning.

References

Archer, E. (2009). Beyond the rhetoric of formative assessment: Seeking solutions for South Africa in New Zealand's Assessment Tools for Teaching and Learning. Unpublished manuscript, University of Pretoria, South Africa.

Arter, J. A. (2001, April 11-13). *Washington assessment professional development program evaluation results.* Paper presented at the Annual Meeting of the National Council on Measurement in Education (NCME), Seattle, WA.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome).* New York: Academic Press.

Brown, G. T. L. (2001). *Reporting Assessment Information to Teachers: Report of Project asTTle Outputs Design.* asTTle Tech. Rep. #15. Auckland, NZ: University of Auckland, Project asTTle.

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: Policy, Principles and Practice*, *11*(3), 305-322.

Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports, 99,* 166-170.

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. (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 Modelling in Educational Research: Concepts and Applications* (pp. 243-266). Rotterdam, NL: SensePublishers.

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

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: Policy, Principles 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.

Croft, A. C., & Reid, N. A. (1991). *How often and for what purposes are NZCER tests used in primary and secondary schools?* (Research Report). Wellington, NZ: NZCER.

Croft, A. C., Strafford, E., & Mapa, L. (2000). *Stocktake/evaluation of existing diagnostic tools in literacy and numeracy in English*. Wellington, NZ: NZCER.

Gipps, C., Brown, M., McCallum, B., & McAlister, S. (1995). *Intuition or evidence? Teachers and national assessment of seven-year-olds.* Buckingham, UK: Open University Press.

Hattie, J. A. (2002). *Schools like mine: Cluster analysis of New Zealand schools.* (asTTle Tech. Rep. No. 14). Auckland, NZ: University of Auckland, Project asTTle.

Hattie, J.A.C. (2006, July). *Large-scale assessment of student competencies.* Paper presented as part of the Working in Today's World of Testing and Measurement: Required Knowledge and Skills, Joint ITC/CPTA Symposium, 26th International Congress of Applied Psychology, Athens, Greece.

Hattie, J. (2009, April). *Visibly learning from reports: The validity of score reports.* Paper presented at the annual meeting of the National Council for Measurement in Education (NCME), San Diego, CA.

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. 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. A. C., Brown, G. T. L., Keegan, P. J., MacKay, A. J., Irving, S. E., Cutforth, S., et al. (2004). *Assessment Tools for Teaching and Learning asTTle) Manual* (Version 4, 2005 ed.). Wellington, NZ: University of Auckland / Ministry of Education / Learning Media.

Hattie, J. A., Brown, G. T. L., Ward, L., Irving, S. E., & Keegan, P. J. (2006). Formative evaluation of an educational assessment technology innovation:

Developers' insights into Assessment Tools for Teaching and Learning (asTTle). *Journal of Multi-Disciplinary Evaluation, 5,* Available online: http://survey.ate.wmich.edu/jmde/index.php/jmde_1/article/view/50/57.

Heaton, J. B. (1975). *Writing English language tests*. London: Longman.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Westport, CT: Praeger.

McDowall, S., Cameron, M., Dingle, R., Gilmore, A., & MacGibbon, L. (2007). *Evaluation of the Literacy Professional Development Project* (RMR No. 869). Wellington, NZ: Ministry of Education, Research Division.

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice, 20*(1), 20-32.

McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research, 95*(4), 203-213.

National Research Council. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

Parr, J. M., & Timperley, H. (2008). Teachers, schools and using evidence: Considerations of preparedness. *Assessment in Education: Policy, Principles and Practice, 15*(1), 57-71.

Parr, J. M., Timperley, H., Reddish, P., Jesson, R., & Adams, R. (2007). *Literacy Professional Development Project: Identifying Effective Teaching and Professional Development Practices for Enhanced Student Learning* (RMR No. 851). Wellington, NZ: Ministry of Education, Research Division.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of Classroom Assessment: Learning, Achievement, and Adjustment* (pp. 53-58). San Diego, CA: Academic Press.

Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders* (6th ed.). Boston: Allyn & Bacon.

Quilter, S. M. (1998). *Inservice teachers' assessment literacy and attitudes toward assessment.* Unpublished doctoral dissertation, University of South Carolina, Columbia, SC.

Scriven, M. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation & education: At quarter century* (Vol. Part II, pp. 19-64). Chicago, IL: NSSE.

Shohamy, E. (2001). *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. Harlow, UK: Pearson Education.

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice, 20*(3), 5-15.

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

Ward, L., Hattie, J. A., & Brown, G. T. (2003, June). *The Evaluation of asTTle in Schools: The Power of Professional Development.* asTTle Tech. Rep. #35, University of Auckland/Ministry of Education.