

No: 030

TEACHER'S ENACTED CURRICULUM: UNDERSTANDING TEACHER BELIEFS AND PRACTICES OF CLASSROOM ASSESSMENT

Gavin T L Brown<sup>1</sup>

*The Hong Kong Institute of Education, Hong Kong*

Lois R. Harris

*The University of Auckland, New Zealand*

Paper presented at the International Association for Educational Assessment annual conference, Bangkok, Thailand, August, 2010

**Acknowledgement:** Funding for this study was provided by the Measuring Teachers' Assessment Practices Project (Faculty of Education, The University of Auckland) and the FRSS Fund (Dept of Curriculum & Instruction, The Hong Kong Institute of Education). The cooperation of teachers in Auckland, New Zealand is greatly appreciated.

Abstract

This paper examines the relationship between teacher conceptions of assessment and the curriculum content and cognitive demand of their classroom assessment practices, working from the assumption that what is assessed should reflect the teacher's understanding of assessment. In this exploratory study, a volunteer sample of New Zealand teachers ( $n=9$ ) provided 32 self-selected samples of assessments they used in the subject English and indicated their purposes and uses as well as their conceptions of assessment. Each assessment was rated for its content and cognitive demand using the American *Surveys of Enacted Curriculum* (SEC) project taxonomy. There were no statistically significant relationships between teacher beliefs about assessment and their enacted practices. The study found that in the main, the assessments required low levels of cognitive demand, focusing on memory, recall, explaining, and following procedures rather than analysis and evaluation. Curriculum and cognitive demand in the assessments varied considerably between teachers. This paper suggests that the SEC taxonomy also needs revisions if it is to be used effectively to code teacher-made assessments, especially those underpinned by the principles of Assessment *for Learning*.

Introduction

The understandings teachers have of curriculum have been found to shape their classroom assessment practices (Calderhead, 1996; Thompson, 1992) and predict the types of assessments they implement. New Zealand primary school teachers who conceived of assessment as being for improvement agreed that they used informal assessment practices more than formal tests, while the belief that assessment was for making students accountable predicted greater use of formal tests with surface levels of cognitive demand (Brown, 2009). Conceptions of subject matter have also been found to shape assessment practices. In a review of international research, different major conceptions of the subject mathematics (i.e., relational understanding vs. instrumental understanding) were claimed to be "at the root of disagreements about what constitutes 'sound' approaches to the teaching of mathematics and what constitutes 'sound' student assessment practices" (Thompson, 1992, p. 133). For example, those who conceived of mathematics in relational terms appeared to emphasise authentic, problem-solving, and process-focused forms of assessment, while those who conceived of mathematics in instrumental terms tended to use correct answer-focused forms of assessment.

To further examine the relationship between teachers' curriculum priorities and their enacted practices, it would be useful to look at the enacted curriculum embedded in the content of their assessments. Under the general assumption of '*what you test is what you get*', it seems logical to presume that the aspects of the curriculum that teachers deem important will get assessed. Of course, if assessment practices are controlled by school or jurisdictional policies, teachers may have little personal influence over the curriculum enacted in their assessments. Hence, research examining this hypothesis would need to take place in a setting where teachers

---

<sup>1</sup> Correspondence: Dr Gavin T L Brown, Dept of Psychological Studies, The Hong Kong Institute of Education, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong SAR; Email: [gtlbrown@ied.edu.hk](mailto:gtlbrown@ied.edu.hk)

generally control the assessment process.

While the relationship between teachers' curriculum priorities and their assessment has been studied, the relationship of teachers' conceptions of the nature and purpose of assessment in general to the content of their classroom assessment practices is less well understood. Brown (2008) identified four major purposes for assessment (i.e., improved learning and teaching, school evaluation, student evaluation, and irrelevance). These four conceptions of assessment have been shown to relate to teacher beliefs about the nature of teaching, learning, and curriculum and to self-reported assessment practices (Brown, 2008, 2009). It seems likely that if teachers conceive of the dominant purpose of assessment to be student evaluation, they would tend to use relatively formal, and possibly standardized, assessment tasks that focus on the surface features of the curriculum most readily assessed in paper-and-pencil formats. In contrast, if improved learning is the dominant purpose, teachers, at least in jurisdictions that do not practice high-stakes national testing, would tend to use informal tasks that focus on higher-order cognitive tasks and complex, integrated performances. Thus, it is expected that teachers' purposes for assessment could be a contributing factor in the curriculum enacted in their assessment practices.

Thus, a goal of this study was to explore the possibility that teachers' beliefs about the goals and purposes of assessment have systematic relationships with the curriculum content and cognitive demand of the assessments they actually use. This study was an exploratory investigation in New Zealand of the assessment practices of nine teachers within the context of the English curriculum. New Zealand was chosen because of the low-stakes way assessment is practiced through the first 10 years of schooling. Within New Zealand, schools are self-governing and the New Zealand Ministry of Education is committed to the use of Assessment *for* Learning (A<sub>f</sub>L), expecting schools to use assessment data to improving student learning. There are no compulsory national assessments or tests in the New Zealand school sector (Crooks, 2002), with school quality judged through a range of measures (e.g., school visits and evaluations conducted by the Education Review Office, National Education Monitoring Project). In the last decade, two significant resource initiatives [i.e., the Assess to Learn (ATOL) professional development program and asTTle, a computer-assisted assessment tool] have been deployed nationally to assist teachers with assessment practices (Brown, Irving, & Keegan, 2008; Hattie, et al., 2004); these focus on using assessment formatively to improve learning. Under the Ministry of Education (1994b) assessment policy, schools are able to select whatever Ministry assessments they believe are valid and appropriate for both reporting student learning and guiding teaching and learning activities in the classroom. Thus, wide usage of different government-funded tools alongside teacher created tasks would be expected to assess student learning of curriculum standards within New Zealand; however, this process is entirely controlled at the school level during years 1-10.

At the time of this study, a new curriculum had been announced (Ministry of Education, 2007) and was being gradually phased in. This English curriculum has two strands which integrate language and literature across three modes of communication. Students are expected to (1) make meaning of ideas or information they receive through listening, reading, and viewing and (2) create meaning for themselves or others by speaking, writing, and presenting. This is substantially the same material, albeit structured differently, as the 1994 English curriculum (Ministry of Education, 1994a). To assess these strands, teachers and schools were allowed to create their own assessments, or draw on a large range of diverse assessments created by the Ministry (details in Brown, Irving, & Keegan, 2008). These tools include more traditional standardized tests, as well as performance tasks and annotated exemplars.

While a major line of investigation within this study related to how teacher conceptions of assessment related to the content and cognitive demand of their enacted assessment process, methodological questions about how the content of these enacted assessments could be appropriately categorized and calculated were also examined. The American Surveys of Enacted Curriculum (SEC) project taxonomy was selected as the framework for coding the curriculum and cognitive content within this study as it has been successfully used to examine the alignment of state standards, assessment tasks, and curriculum content in mathematics and English language arts/reading (Porter, 2002; Porter & Smithson, 2001, 2002; Porter, Polikoff, Zeidner, & Smithson, 2008). In the SEC methodology, assessment tasks are classified at the item level against

achievement objectives and strands provided in the taxonomy for both curriculum content and cognitive demand. In English Language Arts/Reading (ELA/R), the major strands for reading have a hierarchical arrangement (i.e., simpler content has been assigned lower numbers than more complex content; values 100 to 800); however, the remaining tasks do not seem to be inherently more demanding in terms of curriculum than the various reading tasks. Likewise, the cognitive demand categories have a similar hierarchical arrangement (i.e., less complex are coded earlier than more complex).

Of three major alignment methods identified in the literature, the SEC methodology has reported reliability of judges to make similar ratings across instructional and assessment tasks (Martone & Sireci, 2009). While there are obviously differences between the English curriculum in various American states and the national New Zealand English curriculum, because this taxonomy provides a wide range of coding categories and has been found to be reliable in other studies, it was deemed appropriate to trial with New Zealand curriculum. The SEC ELA/R is a relatively young document, having undergone only one revision (J. Smithson, personal communication, 18 December 2009). This means that coding conventions that clarify distinctions between the strands and achievement objectives have not been fully developed. Hence, part of this trial was also designed to examine the utility of the ELA/R taxonomy for classifying both assessments in general and also for specific use with New Zealand English assessment tasks.

#### Method

This study, using a small sample of volunteer cases (Yin, 2009), had a non-experimental design mixing both factor analysed survey responses (Kline, 1994) and content analysis (Smith, 2000) methods. The study was part of the Measuring Teachers Assessment Practices project, which was a multi-faceted, two-year investigation into teacher assessment beliefs and practices carried out in New Zealand.

#### *Procedures*

Initially, over 160 teachers completed the abridged Teachers' Conceptions of Assessment inventory (Brown, 2006) and 26 teachers agreed to be interviewed (Harris & Brown, 2009). In a second round of data collection, out of the 100 teachers who agreed to participate beyond completing the initial survey, 14 primary and secondary teachers agreed to supply a self-selected sample of assessments they had used in either English or mathematics. Teachers were asked to provide six examples of assessment tasks and complete a short questionnaire for each which explained its origins and purposes; primary teachers were encouraged to provide a mix of English and mathematics tasks. While teachers were asked to provide assessments which were representative of their typical practice, it is difficult to claim that in all cases the samples reflect 'typical' assessment practices within the classrooms since only four teachers from this sample were observed in the classroom.

In all, nine teachers provided 32 assessments which they used in English. There were 339 separate identifiable items, tasks, or questions within those assessments. The curriculum and cognitive characteristics of each assessment item in the tasks were classified according to the SEC ELA/R strands. Rating was carried out by the second author, an experienced English teacher who has taught in Australia and the United States. The first author, an experienced New Zealand English language arts teacher, validated this coding and resolved uncertainties as to the meanings of the curriculum categories. Items were classified according to the major strands of the taxonomy (Table 1) and were assigned to multiple strands when applicable.

#### *Participants*

The nine teachers were employed in nine different schools. Two worked in secondary schools (Years 9-13), one at an intermediate school (Years 7-8), and the balance at full (Years 1-8) or contributing (Years 1-6) primary schools. Four teachers provided assessments for Year 6 (nominally aged 10); three for Years 7 and 8, and two for Year 10 (nominally aged 14). Eight of the nine teachers were female. All were of New Zealand European ethnicity. Two were associate, assistant, or deputy principals; two were senior teachers; and the balance was regular classroom teachers.

#### *Instruments*

*Conceptions of Assessment.* The abridged Teachers' Conceptions of Assessment Inventory was used to measure the teachers' conceptions of assessment. It has 27 items which aggregate into four major purposes. These have been shown to have good fit to the data in studies with New Zealand primary and secondary teachers and Queensland primary teachers (Brown, 2007, 2008). For each teacher, their general attitude towards the four purposes of assessment was obtained from the average of their responses to the items associated with each purpose.

Table 1. Frequency of SEC ELA/R Taxonomy Strand Classifications

Curriculum Content	Weight	<i>N</i>	Cognitive Demand	Weight	<i>N</i>
<b>100</b> Phonemic Awareness	100	0	<b>B</b> Memorize/Recall <b>C</b> Perform	100	91
<b>200</b> Phonics	200	5	Procedures/Explain <b>D</b> Generate/ Create/	200	180
<b>300</b> Vocabulary	300	10	Demonstrate	300	62
<b>400</b> Text and print features	400	14	<b>E</b> Analyze/ Investigate	400	32
<b>500</b> Fluency	500	0	<b>F</b> Evaluate	500	4
<b>600</b> Comprehension	600	66			
<b>700</b> Critical reading	700	23			
<b>800</b> Author's craft	800	26			
<b>900</b> Writing process	500	64			
<b>1000</b> Elements of presentation	800	70			
<b>1100</b> Written applications	600	121			
<b>1200</b> Language study	600	30			
<b>1300</b> Listening and Viewing	600	7			
<b>1400</b> Speaking and presenting	600	24			
<i>Total</i>		460			438

*Assessment tasks.* Teachers were asked to complete a short questionnaire to accompany each assessment item. Teachers provided data about the sources of the assessment (e.g., teacher made, made by colleague, externally created) and the reasons had they had for using it. A total of 27 different reasons were listed and teachers were asked to indicate all that applied. Specifically, the reasons were aggregated into five major categories (i.e., administration, assessment for learning (AfL), evaluation, planning, and reporting), which were used for analytic purposes. The AfL category was based on the Ministry of Education's (2007) policy priority, which meant this category included purposes relating to student engagement and motivation (e.g., making work fun and interesting) alongside student-centered assessment practices (e.g., self- and peer-assessment, students setting learning targets); it is noted that the reasons relating to student engagement and motivation may not actually lead to greater learning (Harris, 2008). Generally, the reasons fell into improvement (i.e., AfL and planning) and accountability (i.e., administration, evaluation, and reporting) oriented reasons.

## Results

### *Conceptions of Assessment*

Through responses to the *Teachers' Conceptions of Assessment* inventory, the nine teachers agreed that assessment evaluates school quality more than any other conception (Cohen's  $d > 1.60$  [Cohen, 1992]) (Table 2).

Previous studies with New Zealand teachers had found that school accountability was one of the least endorsed purposes of assessment (Brown, 2007, 2008). Brown and Harris (2009) suggested that the strong emphasis on this purpose has arisen in part because of schooling improvement initiatives in which school-wide assessment results are used to fulfill school self-

and external-review expectations. It is also noteworthy that the correlation between student accountability and improvement was much larger than previously found among New Zealand teachers and approached values seen in a Hong Kong study with the same instrument (Brown, Kennedy, Fok, Chan, & Yu, 2009). This demonstrated that this small group of teachers conceived of assessment as demonstrating school quality and that grading or evaluating students was associated with improved teaching and learning.

Table 2. Teachers' Conceptions of Assessment inventory factor statistics

Statistic	TCoA Factors			
	I.	II.	III.	IV.
<i>M</i>	3.67	4.81	3.60	3.31
<i>SD</i>	0.62	0.63	0.38	0.30
	Pearson ( <i>r</i> )		Mean score differences (Cohen's <i>d</i> )	
I. Student Accountability	—	<b>-1.84</b>	<b>0.13</b>	<b>0.76</b>
II. School Accountability	<i>.11</i>	—	<b>2.40</b>	<b>3.25</b>
III. Improvement	<i>.80</i>	<i>.32</i>	—	<b>0.84</b>
IV. Irrelevance	<i>.25</i>	<i>-.32</i>	<i>.11</i>	—

Note. Pearson correlations below diagonal in italics; Cohen's *d* effect size above diagonal in bold.

#### *Assessment Tasks*

Of the 32 tasks, 13 were writing or editing, 9 were reading comprehension, three were visual text analyses, four were spoken language performance, and three were grammar or spelling activities. Hence, the priority in terms of curriculum coverage was written language (both reading comprehension and writing) and relatively equal between making and communicating meaning. It is unknown whether this is a reflection of the medium of assessment (paper-and-pencil) or a true reflection of the curriculum priorities. Over two-thirds ( $n=23$ , 72%) of the 32 assessments were created by the teachers or their colleagues, with the balance coming from external sources. Of those sources, asTTle, national exemplars, Progressive Achievement Tests, and the Assessment Resource Banks were all used more than once.

A total of 247 different responses were given to the 28 specific purposes for using the assessments. Barring the five 'other' responses, the 247 responses were dominated by three major categories: reporting ( $n=66$ , 27%), evaluation ( $n=55$ , 22%), and assessment *for* learning ( $n=74$ , 30%). Planning ( $n=44$ , 18%) and administration ( $n=3$ , 1%) comprised the balance of responses. Hence, real educational functions around improvement and accountability are given for using these assessments, with accountability and improvement purposes being nearly equal (i.e., reporting, evaluation & administration=50%; assessment for learning & planning=48%, with 2% being 'other').

#### *Curriculum Characteristics*

A total of 460 ratings were made for curriculum strand and 438 ratings were made for cognitive demand (Table 1). Only 31% of tasks were related to making meaning by reading text. It is clear that a large portion of tasks focused on writing applications, with a high emphasis on the cognitive processes: memorize, recall, perform procedures, and explain.

In order to examine relationships between conceptions of assessment and the SEC ELA/R ratings, a weighted profile for each of the 32 tasks for both curriculum content and cognitive demand was created. Close inspection of the tasks provided suggested that most teachers were implementing the less demanding aspects of these categories which led to the adoption of the weighting system shown in Table 1 in which a weight was assigned to the writing, listening, and speaking tasks identical to the reading task that was judged to be the most similar in demand. Since some assessment tasks had many more items than others (e.g., the standardized tests), a weighted score for each task was created and the average for the teacher was found for the two to six tasks each had provided.

The resulting weighted scores show the relative importance of the curriculum content and cognitive demand for each teacher (Table 3). Teachers with higher mean scores provided assessments that require more demanding content and/or cognitive performances. Note that only three teachers had  $SD < 10\%$  of the mean value for curriculum and one for cognitive demand.

Hence, the assessment tasks were highly variable in what was demanded for curriculum and cognition.

Table 3. Teacher profiles for enacted curriculum and cognitive demand and conceptions of assessment.

Teacher	<u>Weighted SEC ELA/R</u>		<u>Teacher Conceptions of Assessment</u>			
	Curriculum	Cognitive	Student Accountability	School Accountability	Improvement	Irrelevance
T1	624.44 (65.52)	230.43 (31.66)	3.00	4.00	3.00	3.33
T2	666.07 (48.96)	238.89 (53.44)	3.00	4.33	3.58	3.50
T3	597.78 (91.42)	164.81 (107.82)	3.33	5.33	3.25	2.89
T4	498.08 (217.10)	224.40 (21.43)	4.33	5.67	4.08	3.67
T5	663.89 (94.58)	186.20 (180.00)	3.33	5.67	3.58	3.11
T6	651.16 (130.98)	192.40 (80.84)	4.33	4.67	3.92	3.22
T7	678.33 (58.37)	268.20 (63.03)	3.33	5.00	3.67	3.00
T8	648.57 (12.12)	251.10 (40.86)	3.67	4.33	3.25	3.78
T9	589.29 (65.34)	173.10 (50.95)	4.67	4.33	4.08	3.33
<i>M</i>	624.18	214.41	3.67	4.81	3.60	3.31
<i>SD</i>	87.15	51.34	0.62	0.63	0.38	0.30
<u>Pearson Correlations</u>						
Curriculum		0.28	-0.54	-0.34	-0.38	-0.32
Cognitive			-0.38	-0.28	-0.21	0.42

Note that none of the correlations were statistically significant because  $r < .67$ . Hence, while it may be possible to interpret the observed values, it is best to conclude there is no systematic relationship between how teachers conceive of the competing purposes of assessment and their enacted curriculum choices in the assessment tasks provided. Clearly, a much larger sample is required to establish with statistical significance the nature of these relationships.

#### Discussion

In general, the assessment tasks exhibited relatively low levels of cognitive demand (i.e., recalling, memorising, following procedures, and explaining) as opposed to tasks which required analysis and evaluation. The mean curriculum score was 624.18, indicating a dominance of tasks which required comprehension or basic text production with little critical reading and consideration of audience and purpose in writing. This curriculum content may be expected at these grade levels as this content is considered foundational; however, it raises concerns that, if assessments do not require higher order curriculum tasks, students will not acquire those skills. It was interesting to note the dominance of reading and writing tasks as opposed to those focused on speaking and viewing; if these domains are considered to be of equal importance within the curriculum, clearly a shift is needed to produce more emphasis on speaking and viewing tasks. However, these results may be an artifact of teachers not providing assessments for higher order curriculum and cognitive challenges in English that they consider cannot be legitimately evaluated through paper-and-pencil assessments (e.g., informal and performance assessments).

While it would be tempting to interpret the statistically non-significant correlations between conceptions of assessment and curriculum and cognitive demand, this will not be discussed. What is more apparent is that factors other than the personal conceptions of the nature

of assessment must have influence on teacher's curriculum choices in their assessment tasks. It seems probable that the official English curriculum and school-level curriculum or assessment priorities play a significant role. Future research into the enacted curriculum of teacher-controlled assessments should focus more on their curriculum beliefs.

While small sample size obviously affected the power of these results, problems were also encountered when using the SEC framework for evaluating New Zealand English assessment items. Some classification caused conflicting responses as they had similar wording (e.g., word meaning or vocabulary occurred in strands 300 and 600, inference occurred in strands 600 and 700, spelling occurred in 900 and 1200). As the taxonomy has not been published with a training guide to clarify these cross-overs, it was very difficult to determine what these distinctions might actually refer to. Furthermore, there were no clear guidelines for distinguishing between and among elements of verbal and written presentation (strand 1000), what a writer has to do (strands 900 and 1100), comprehension of material (strand 600), critical reading of texts (strand 700), author's craft analysis (strand 800), listening and viewing of texts (strand 1300) and speaking and presenting (strand 1400). A strategy and system for resolving overlap between strands is needed.

Smithson (personal communication, 18 December 2009) indicated that reference to the more global or coarse grain curriculum construct would be used to resolve these category distinctions. This would require consensus among local curriculum experts to develop rules by which distinctions could be made within a curriculum jurisdiction. This would increase the validity of the taxonomy for usage in any one jurisdiction, while reducing the utility of the taxonomy for usage beyond the context where these judgments were made.

An additional problem emerged when trying to use these categories to mathematically calculate weights and scores for tasks. First, within each strand (i.e., 100, 200, 300, etc.) there were often items which greatly varied in difficulty. The weightings we used reflected the way we saw these categories enacted in our particular data set; however, these same weightings would not be valid if teachers had been using more complex curriculum content under certain categories. For example in 1200, Language Study, our teachers were looking for identification of simple parts of speech rather than the higher level linguistic analyses described later in the category. How to mitigate the broad differences in demand within categories remains unresolved.

Finally, facets of the New Zealand English curriculum were not well-handled by the taxonomy. For example, category 400 text and print features relates to printed texts. However, the English curriculum in New Zealand requires literary and technical analysis of static and moving images which are treated as texts which may belong to strand 1300 rather than 400. Furthermore, research skills are considered part of the English curriculum in New Zealand and were assessed in this sample, but these do not appear explicitly in the taxonomy. We agree with Smithson (personal communication, 18 December 2009) that the development of a new category for research skills should be implemented in future versions of the taxonomy. Additionally, in an assessment context with an AfL focus, student self-assessment tasks were relatively common; where these might fit against curriculum and cognitive levels was often unclear. Again, it would appear that reference to the global context and allowing for multiple categorization of each task is needed to fit New Zealand's English curriculum assessments to the SEC taxonomy. Hence, in its current format the SEC taxonomy would appear to have limited utility for application beyond a US type ELA/Reading curriculum.

#### References

- 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 to the NZARE annual conference, December, 2007, 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 modeling in educational research: Concepts and applications* (pp. 243-266). Rotterdam, NL: Sense Publishers.
- 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., Irving, S. E., & Keegan, P. J. (2008). *An Introduction to Educational Assessment, Measurement, and Evaluation: Improving the Quality of Teacher-Based Assessment*. (2nd ed.). Auckland, NZ: Pearson Education NZ.
- 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 & Practice*, 16(3), 347-363.
- Calderhead, J. (1996). Teachers: Beliefs and knowledge. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 709-725). New York: Simon & Schuster Macmillan.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Crooks, T. J. (2002). Educational assessment in New Zealand schools. *Assessment in Education: Principles, Policy & Practice*, 9(2), 237-253.
- Harris, L. R. (2008). A phenomenographic investigation of teacher conceptions of engagement in learning. *Australian Educational Researcher*, 35(1), 57-79.
- Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy and Practice*, 16(3), 365-381.
- Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.
- Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79(4), 1332-1361.
- Ministry of Education. (1994a). *English in the New Zealand curriculum*. Wellington, NZ: Learning Media.
- Ministry of Education. (1994b). *Assessment: Policy to practice*. Wellington, NZ: Learning Media.
- Ministry of Education. (2007). *The New Zealand Curriculum for English-medium teaching and learning in years 1-13*. Wellington, NZ: Learning Media.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.
- Porter, A. C., & Smithson, J. L. (2001). *Defining, developing, and using curriculum indicators* (CPRE Research Report No. RR-048). Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.
- Porter, A. C., & Smithson, J. L. (2002, March). *Alignment of assessments, standards and instruction using curriculum indicator data*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Porter, A. C., Polikoff, M. S., Zeidner, T., & Smithson, J. (2008). The quality of content analyses of state student achievement tests and content standards. *Educational Measurement: Issues & Practice*, 27(4), 2-14.
- Smith, C. P. (2000). Content analysis and narrative analysis. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 313-335). Cambridge, UK: Cambridge University Press.
- 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.
- Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Los Angeles: Sage Publications.