Item Signature Study: Report on the Characteristics of Reading Texts and Items from Calibration 3

Technical Report 28: Project asTTle, March 2002

Gavin T. L. Brown
University of Auckland

This is the report of the third item signature study conducted with Project asTTle reading assessment items. Eight practising teachers determined the cognitive and curriculum characteristics of 76 assessment items and the curriculum and structural characteristics of 10 reading texts. Ratings were conducted in pairs, providing excellent dependability (φ>0.80) and efficiency.

Table of Contents

Introduction

The Assessment Tools for Teaching and Learning Project (Project asTTle), a partnership between the University of Auckland and the Educational Testing Centre, University of NSW, Australia, is developing, under contract to the Ministry of Education, new literacy and numeracy assessment tools for use with students in Years 5 to 7, in English and in Maori.

These tools, to be supplied on a CD-ROM, will extend the range of voluntary-use assessment tools currently available to primary schools. Using the asTTle item "banks" of reading assessment items (or tasks) on the CD-ROM, schools will be able to analyse and aggregate information about the achievement of individual students and/or groups of students within the school in relation to national standards at any time during the year. All assessment items in the banks have been initiated and reviewed by New Zealand teachers and have been calibrated on a nationally representative sample of students.

This report should be read in conjunction with two previous item signature studies that ascertained the literacy-relevant characteristics of other items and texts in the asTTle item bank (Meagher-Lundberg & Brown, 2001a; Meagher-Lundberg & Brown, 2001b). These characteristics will be specified by test users (usually teachers) in compiling tests from assessment items in the bank using the program on the CD-ROM (Brown, 2001).

The item signature process of determining assessment item characteristics was documented by Burstein, Koretz, Linn, Sugrue, Novak, Baker, & Harris (1995/1996). Using a similar method to that used in the second item signature study, this study ascertained the characteristics of the reading texts and assessment items calibrated in November 2001 for the asTTle item bank. The third calibration consisted of items and texts designed to assess close reading at Level 4 of the English curriculum. This highly focused calibration was conducted in order to ensure sufficient materials were available in Level 4.

Purpose of the Study

The purpose of the study was to critically classify reading texts and assessment items from the third calibration – Papers I to K – for the Close Reading strand of the English

curriculum, to ensure they met quality standards.

Goals of the Study

The goals of the study were:

1. to rate the reading texts and assessment items from Papers I to K according to the definitions. and
2. to obtain recommendations about future assessment item writing design specifications.

Methodology

Definition Development

Definitions for key literacy categories developed for Item Signature Study 2 (Meagher-Lundberg & Brown, 2001b), definitions were used in this study. The literacy categories covered both text features and comprehension processes (i.e., the processes involved in answering an assessment item or task). Definitions were developed from English curriculum documents, international assessment programmes, and reports of various research projects conducted for Project asTTle.

Categories and Variables used to Classify Reading Texts

The ten reading texts were rated according to the literacy categories listed below. Each category has two or more variables related to the features of a reading text. The variables are mutually exclusive – that is, only one variable under each category is applicable to any one reading text. A brief definition of each category is given below.

- *Form* – The shape of the text in either continuous or non-continuous form affects the nature of communication. This category is based on the PISA classification system (Kirsch, Mendelovits, & McQueen, 2000).
- *Curriculum Level* – Levels of achievement in the New Zealand English curriculum (Ministry of Education, 1994).
- *Difficulty within Curriculum Level* (rated in conjunction with curriculum level) – The level of difficulty of text within the assigned curriculum level (Ministry of Education, 1994; NCES, 2001).
- *Purpose* – There are two major purposes of texts adopted from the IEA PIRLS study: literary or informational (Campbell, Kelly, Mullis, Martin, & Sainsbury, 2001).
- *Print Considerateness* – The level of assistance the text offers readers in understanding the meaning.
- *Illustration Considerateness* – The level of assistance illustrations offer to the reader in understanding the text.
- *Genre (Purpose)* – Within each purpose there are a variety of purposes and the intended audiences that further distinguish texts (Glasswell, Parr, & Aikman, 2001).
- *Essential Learning Area* – The content or topic of each text can be described by the Essential Learning Area (defined by the New Zealand Curriculum Framework) that it belongs to (Ministry of Education, 1993).

Categories and Variables used to Classify Assessment Items

The 76 assessment items were classified according to the categories listed below. These relate to the comprehension and cognitive processes used in answering an assessment item (or task) and to curriculum objectives of that item. Each category has two or more variables, which are mutually exclusive – that is, only one variable under each category is applicable to any assessment item – with the exception of Curriculum Area Objectives, for which any number of objectives may be applicable. A brief definition of each category is given below.

- *SOLO Taxonomy* – Depth of cognitive processing as defined the Structure of Observed Learning Outcomes (SOLO) taxonomy (Biggs & Collis, 1982; Hattie, & Purdie, 1998).
- *PIRLS Processes of Comprehension* – Four comprehension processes that enable students to construct meaning from written texts, taken from the Performance in International Reading Literacy Survey (PIRLS), a study of reading among nine-year-olds (Campbell et al., 2001).

- *Curriculum Area Objectives* – These are the objectives that further specify the achievement aims of the close reading strand of the English curriculum. Curriculum Area Objectives included in this study comprised deep and surface features. The deep features were: find information; knowledge; understanding; connections; and inference. The surface features were: grammar; punctuation; and spelling (Limbrick, Keenan, & Girven, 2001; Ministry of Education, 1994).
- *Curriculum Area Processes* – The English curriculum identifies three cognitive processes integral to the development of literacy (Ministry of Education, 1994).

All 76 assessment items included in this study were linked to a reading text. Rating responses were mutually exclusive for SOLO Taxonomy, PIRLS Processes of Comprehension, and Curriculum Area Processes. Curriculum Area Objectives were single-value choices, with raters assigning as many objectives as they considered applicable to each assessment item.

Workshop and Quality Control Processes

A two-day workshop was run to critically classify, or rate, the reading texts and assessment items from Papers I to K according to the definitions provided for and refined within the workshop. Training procedures were designed around those documented by Baker, Aschbacher, Niemi, & Sato (1992). One reading text from Paper I and its 7 assessment items were used for training in a whole group consensus setting. These items were then included as part of this report.

Eight teachers – all with English curriculum and teaching experience, and previous experience of the classification work from the previous item signature studies – took part in the workshop. The presenter for the workshop, the author of this report, has been involved in all previous item signature studies and is a member of the Project asTTle development team.

The key task required of participants in the workshop was to rate the reading texts and assessment items according to the categories and associated variables outlined above. Following recommendations based on the outcomes of the second item signature study, a particular emphasis of this study was the use of pair rating technique. All rating decisions were made on the basis of three out of four pairs assigning a rating to an item or text. Where that level of agreement was not reached group discussion took place to reach consensus.

The morning of Day One focused on refreshing teacher understanding of the definitions for rating reading texts and completing the ratings of texts. The balance of the workshop focused on understanding and using the definitions for rating assessment items.

Rating procedure. The procedure devised for the teachers to rate the reading texts and assessment items was as follows: (a) discussion of definitions as a whole group and rating of the training texts/items as a group to ensure group understanding of definitions (calibration); (b) rating of texts/items by pairs of teachers; and (c) where there was disagreement (i.e., where three-quarters majority agreement was not reached) on the rating of a text/item in one or more category, group discussion of definitions; followed until a consensus was reached.

When rating texts and items, teachers handed in their pair score sheets, which were then tallied. Where there was not a three-quarters majority agreement on a characteristic, a group discussion – involving checking definitions and reviewing previous benchmark decisions – took place until a consensus was reached.

It was agreed that working in pairs, as recommended in the second item signature study, led to high levels of agreement. The dependability of the ratings of assessment items, calculated using the Brennan and Kane Dependability Index (φ)[1], is reported in Table 13 below.

[1] For details see Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory. A Primer*. Newbury Park, CA: Sage.

Findings

Characteristics of Reading Texts

The characteristics of the ten reading texts in this item signature study are presented below.

Text Form

All of the ten reading texts were characterised as continuous that is, composed of sentences and organised into paragraphs.

Text Purpose (PIRLS)

Of the ten reading texts in this study, almost all (90%) were assigned the characteristic *acquire and use information* and only 1 was deemed to provide a *literary experience* (Table 1).

Table 1
Characteristics of Text Purpose (PIRLS)

Text Purpose (PIRLS)	Number	% of Total
Information	9	90%
Literary experience	1	10%
Total	10	100%

Note. Texts are deemed to have a characteristic where there is a three-quarters agreement or better.

Curriculum Level and Difficulty

Table 2 shows the assignment of reading texts to Curriculum Level and Difficulty with Level. Almost all (90%) of the reading texts were assigned to Level 4 of the curriculum, with one text (10%) assigned to Level 5. Four of the reading texts were classified as 4 proficient, and another four were as 4 advanced.

Table 2
Text Characteristics by Curriculum Level and Difficulty

	Level 4		Level 5	
Curriculum Difficulty	Number	% of total	Number	% of total
Basic	1	10%	1	10%
Proficient	4	40%	0	–
Advanced	4	40%	0	–
Total	9	90%	1	10%

Note. Texts are deemed to have a characteristic where there is a three-quarters agreement or better.

Print Considerateness

All reading texts were characterised as having considerate print; that is, print that helps readers to comprehend the content of the texts.

Illustration Considerateness

Considerate illustrations assist readers to comprehend the content of the reading text. Half of the ten reading texts were not applicable to this category, as they had no illustrations. The remaining five texts were all characterised as having considerate illustrations (Table 3).

Table 3
Characteristics of Illustration Considerateness

Illustration Considerateness	Number	% of Total Applicable
Considerateness of Illustration	5	50%
Non-considerateness of Illustration	0	–
Not Applicable	5	50%
Total	10	100%

Note. Texts are deemed to have a characteristic where there is a three-quarters agreement or better.

Genre (Purpose)

The term genre is used to describe the range of processes (entertaining, reporting, instructing, arguing or persuading, explaining, and informing) used to produce texts that reflect a purpose and audience. The purpose of the text, not its form, determines the genre; for example, a letter can explain or argue.

Nearly three-quarters (60%) of all reading texts were characterised as *persuading* and nearly one-third (30%) as *explaining* (Table 4).

Table 4
Characteristics of Genre (Purpose)

Genre (Purpose)	Number	% of Total
Argue	6	60%
Explain	3	30%
Entertain	1	10%
Report	-	-%
Instruct	-	-%
Inform	-	-%
Total	10	100%[a]

Note. Texts are deemed to have a characteristic where there is a three-quarters agreement or better.

Essential Learning Area

The New Zealand Curriculum specifies seven essential learning areas that describe in broad terms the knowledge and understanding that all students need to acquire and incorporate during the first ten years of school. Five of the seven essential learning areas were considered applicable to the 10 reading texts. Mathematics and Technology were not assigned to any text.

As Table 5 shows, 4 (40%) of the ten reading texts were characterised as relevant to Language. Three (30%) were characterised as relevant to Science, and just one each to Social Sciences, The Arts, and Health.

Table 5
Characteristics of Essential Learning Area

Essential Learning Area	Number	% of Total
Language	4	40%
Science	3	30%
Social Sciences	1	10%
The Arts	1	10%
Health	1	10%
Technology	-	-
Mathematics	-	–
Total	10	100%[a]

Note. Texts are deemed to have a characteristic where there is a three-quarters agreement or better.

Characteristics of Assessment Items

The characteristics of the 76 assessment items in this study are presented below. Items are listed in order of Paper.

SOLO Taxonomy

For the SOLO cognitive processing taxonomy, the largest proportions of the 76 assessment items were characterised as either relational (59%) or unistructural (21%) (Table 6). Multistructural items made up 16% of the total, while just 4% of items were rated as extended abstract. Surface items (i.e., unistructural and multistructural) accounted for just over one-third (37%) of the items and deep items (i.e., relational and extended abstract) made up the balance (63%).

Table 6
Characteristics of Items by SOLO Taxonomy

SOLO Taxonomy	Number	% of Total
Unistructural	16	21%
Multistructural	12	16%
Relational	45	59%
Extended Abstract	3	4%
Total	76	100%

Note. Items are deemed to have a characteristic where there is a three-quarters agreement or better.

Processes of Comprehension

There are four comprehension processes that enable students to construct meaning from written texts (Table 7). These are: focusing on and retrieving explicitly stated *information*; making straightforward *inferences*; *interpreting* and integrating ideas and information; and examining and *evaluating* content, language, and textual elements. The largest proportions of assessment items were characterised as either *inference* (37%) or *interpret* (36%) (Table 7). Around a quarter (26%) were characterised as *information*. *Evaluation* was assigned to only one assessment item.

Table 7
Characteristics of Items by Processes of Comprehension

Processes of Comprehension	Number	% of Total
Information	20	26%
Inference	28	37%
Interpret	27	36%
Evaluation	1	1%
Total	76	100%

Note. Items are deemed to have a characteristic where there is a three-quarters agreement or better.

Curriculum Area Processes

There are three curriculum area processes. These require the student to: (a) *explore* how the English language is constructed; (b) *think critically* about language and meaning and develop the skills of literary criticism; and (c) use processes by which *information* is identified, understood, stored, organised, retrieved, combined, and communicated. Over half (55%) of all assessment items were characterised as *thinking critically*. *Processing*

information was assigned to a quarter of the items and *exploring language* to just over a fifth (Table 8).

Table 8
Characteristics of Items by Curriculum Area Processes

Curriculum Area Processes	Number	% of Total
Thinking Critically	42	55%
Processing Information	18	24%
Exploring Language	16	21%
Total	76	100%

Note. Items are deemed to have a characteristic where there is a three-quarters agreement or better.

Curriculum Area Objectives

Unlike characteristics for all other categories, Curriculum Area Objectives were not mutually exclusive. Raters assigned as many objectives as they considered relevant to each of the 76 assessment items (Table 9).

Curriculum Area Objectives comprise deep feature and surface feature objectives. In this item signature study, deep feature curriculum area objectives were grouped under the headings: Find Information; Knowledge; Understanding; Connections; and Inference. Surface feature objectives were grouped under the headings: Grammar; Punctuation; and Spelling.

Assignment of the surface feature objectives is shown in Table 9. Objective (6a) *identify word classes* was assigned to five percent of the assessment items, while a similar percentage were assigned to use of punctuation conventions. As per the design brief of creating hard items, only seven items were identified by two of the five surface objectives.

Of the deep feature objectives that were assigned to the 76 assessment items, between three and five of the various objectives for each curriculum area were used (Table10). The two objectives most commonly assigned to assessment items were deep feature objectives under the Understanding and Inference headings. They were: (3a) *consistently read for meaning* (Understanding), 49% of items and (5b) *make inferences from text* (Inference), 49% of items. A further two deep feature objectives, one pertaining to Knowledge – (2b) *use and understand vocabulary* – and the other to Finding Information – (1b) *find, select, and retrieve information* – were assigned to between a quarter and a third of all assessment items (25% and 37% respectively). About one-fifth of assessment items were categorised as (5a) *explore author intent and purpose* (Inferences), 21% of assessment items, and (1c) *skim/scan for information* (Finding Information), 21%. Other deep feature objectives were assigned to only one to ten items.

With so few items in this study, there were many objectives for which no items were written, so these are not reported.

Table 9
Curriculum Area Objectives – Surface Features Assigned to Assessment Items

Surface Features	
6 Grammar	7 Punctuation
(6a) Identify word classes	(7a) Use appropriately a variety of punctuation conventions: commas, full stops, capital letters, exclamation marks, question marks, quotation marks, brackets
4 items, (5%)	3 items (4%)

Note. Figures show the numbers and percentages of the 76 items that were assigned to each objective. Objectives were not mutually exclusive. Items are deemed to have a characteristic where there is a three-quarters agreement or better.

Table 10
Curriculum Area Objectives – Deep Features Assigned to Assessment Items

Deep Features				
1 Find Information	2 Knowledge	3 Understanding	4 Connections	5 Inference
(1b) Find, select, and retrieve information (28 items, 37%) (1c) Skim/scan for information (17 items, 22%) (1e) Use dictionary, thesaurus, atlas (2 items, 3%)	(2b) Use and understand vocabulary (19 items, 25%) (2c) Use and understand poetic and figurative language use (e.g., rhyme and metaphor) (2 items, 3%) (2f) Use and understand text publishing conventions (3 items, 4%)	(3a) Consistently read for meaning (37 items, 49%) (3b) Identify main ideas in texts (10 items, 13%) (3c) Provide detail to support main ideas (1 item, 1%) (3d) Use understandings and information gained from texts (10 items, 13%)	(4a) Compare similarities and differences both within and between texts (5 items, 7%) (4b) Make links between aspects of text (4 items, 5%) (4c) Make use of prior knowledge (script implicit) (1 items, 1%) (4d) Understand and organise material in appropriate sequences (2 items, 3%) (4f) Make links between verbal and visual information (1 items, 1%)	(5a) Explore author's purpose and question author's intention (16 items, 21%) (5b) Make inferences from texts (37 items, 49%) (5c) Read critically a range of texts for bias, stereotyping, and propaganda (1 item, 1%) (5e) Identify and discuss purposes of text types (1 item, 1%)
47 items (62%)	24 items (32%)	58 items (76%)	13 items (17%)	55 items (72%)

Note. Figures show the numbers and percentages of the 76 items that were assigned to each objective. Objectives were not mutually exclusive.
Items are deemed to have a characteristic where there is a three-quarters agreement or better.

Level of Agreement

Agreement on Texts

Agreement on the assigning of text was relatively good with teachers not reaching agreement on only a few texts for purpose, illustration considerateness, and essential learning area. It was clearly evident that assigning Curriculum Level caused the greatest difficulty (Table 11), such that the Brennan-Kane dependability index only reached .623 (Table 16).

Table 11
Number of Texts on which No Agreement was Reached by Category

		Number of Texts on which No Agreement was Reached on Characteristic						
	Total Number of Texts	Form	Curriculum Level	Purpose (PIRLS)	Text Consider-ateness	Illustration Consider-ateness	Genre (Purpose)	Essential Learning Area
Total	10	0	5	1	0	0	5	2

Note. Agreement was reached when there was a three-quarters agreement or better.

Agreement on Assessment Items

Because of the experience of the second item signature (Lundberg & Brown, 2001b), it was decided to conduct all rating in pairs, except for the training phase. Agreement between pairs was reasonable (Table 12), with slightly more items not agreed upon for the PIRLS comprehension processes (i.e., 21 or 28% of items).

Rater dependability. For the 69 assessment items from Papers I to K where rating was undertaken either in pairs (note the first 7 items from Paper I were rated in the whole group condition), the dependability of rater scoring

was measured using the Brennan and Kane Dependability Index.

The Brennan and Kane Dependability Index (φ) is calculated by obtaining the *between-subjects effects error mean square* and dividing it by the sum of the *absolute error variance of the set of ratings* and itself:

$\varphi = \sigma^2_p / (\sigma^2_p + \sigma^2_{ABS})$

Values greater than .80 are considered dependable. the dependability of these ratings is quite acceptable (φ = .87 to .89) (Table 13 for full calculations).

Overall then, all the ratings for SOLO Taxonomy, PIRLS Processes of Comprehension, and Curriculum Area Processes were dependable, as the values for the dependability index exceeded the .80 threshold, for both the individual and the pair rating conditions.

Table 12
Number of Assessment Items by Category where No Agreement was Reached

		Number of Items on which No Agreement was Reached on Characteristic			
Rating Condition	Total Number of Items	SOLO Taxonomy No. of Items	PIRLS Processes of Comprehension No. of Items	Curriculum Area Processes No. of Items	Curriculum Area Objectives No. of Items[a]
Pair condition	76	14	21	16	13

Note. Agreement was deemed to be reached when there was a three-quarters agreement or better.
[a] Indicates the number of items where there was no objective was agreed on. Curriculum area objectives are not mutually exclusive; therefore the dependability of ratings is not calculated.

Table 13
Calculation of the Brennan and Kane Dependability Index (φ) – Results of the Calculations

Category	No. of Pairs	No. of Items	σ^2_i	σ^2_p	$\sigma^2_{pi,e}$	σ^2_{ABS}	φ
Items - SOLO	4	67	2.168	0.376	1.745	0.058	0.866
Items - PIRLS	4	67	1.774	0.245	0.323	0.031	0.887
Items - Processes	4	67	1.048	0.165	0.332	0.021	0.889
Texts - Levels	4	10	3.211	0.678	0.900	0.411	0.623

Note. Of the total 75 items, 67were rated in pairs. The remaining 8 items were rated by the whole group; hence, the dependability of these ratings is not calculated.

Summary of Findings

Characteristics of the Reading Texts

Ten reading texts were classified.

- *Print Considerateness and Illustration Considerateness* – All ten reading texts were rated as being print considerate, and all five texts that were illustrated were rated as having considerate illustrations.
- *Text Form* – All of the texts were classified as continuous (i.e., composed of sentences and organised into paragraphs).
- *Purpose (PIRLS)* – Nine of the 10 were classified as requiring the reader to *acquire and use information,* while the remaining one provided a *literary experience*.
- *Genre (Purpose)* – Six of the texts had the purpose of *arguing* or persuading the reader, with three classified as *explaining,* and one text *entertained*.
- *Essential Learning Area* – Four of the texts (40%) focused on the Essential Learning Area of Language. Science was covered by three of the reading texts, while other areas covered were Social Sciences, the Arts, and Health.
- *Curriculum Level and Difficulty with Curriculum Level* – Nine of the reading texts were assigned to Level 4 of the curriculum, and one text was assigned to Level 5. As per

the design brief, nine of the texts were at or beyond Level 4 Proficient.

Characteristics of Assessment Items

76 reading assessment items were classified.

- *SOLO Taxonomy* – As per the design brief, nearly two-thirds of items (63%) were deep involving relational or extended abstract cognitive processes.
- *PIRLS Processes of Comprehension* – Approximately one third of the assessment items were characterised as one of retrieving explicitly stated *information,* making *inferences* or *interpreting*..
- *Curriculum Area Processes* – Just over half (55%) of the items were characterised as *thinking critically*.
- *Curriculum Area Objectives* – As per the design brief of creating difficult Level 4 items, the deep feature objectives that were most commonly assigned to assessment items were in the areas of understanding and inference.

Level of Agreement

The key quality control measure of ongoing monitoring of rater agreement ensured that contentious issues were discussed and clarified. The iterative process of monitoring, ongoing discussion, checking of definitions, and review of previous decisions ensured that the raters were able to reach consensus during the course of the workshop. Thus, agreement resolution by a group of literacy experts, as in the first item signature study, was not necessary.

The results of the assessment item rating suggest that the level of consensus were satisfactory when raters work as pairs; Brennan and Kane Dependability Index exceeded the .80 threshold. Nevertheless, further work is required on specifying the curriculum level characteristics of reading passages.

Recommendations for Future Studies

It is recommended that future item writing workshops incorporate the item signature study process as part of the development and review of items and texts. It is also recommended that the item signature item characteristics be used to set item writing specifications for items or texts required in the asTTle item bank. It is especially noted that the intersection of the Solo taxonomy deep features, the PIRLS interpretive and evaluative characteristics, and the curriculum processes of exploring language and thinking critically be emphasised in the writing of difficult items for any reading text. It is also recommended that further research be conducted on establishing curriculum level characteristics for reading passages.

References

Baker, E. L., Aschbacher, P. R., Niemi, D., & Sato, E. (1992). *CRESST performance assessment models: Assessing content area explanations*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Biggs, J. B., & Collis, K. F. (1982) *Evaluating the Quality of Learning – The SOLO Taxonomy*. New York: Academic Press

Burstein, L., Koretz, D., Linn, R., Sugrue, B, Novak, J., Baker, E. L., & Harris, E. L. (1995/1996). Describing performance standards: Validity of the 1992 National Assessment of Educational Progress achievement level descriptors as characterizations of mathematics performance. *Educational Assessment 3*(1), 9–51.

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* New York: McKay.

Brown, G. T. L. (2001). *Reporting Assessment Information to Teachers: Report of project asTTle output designs.* Technical Report 15. University of Auckland, Project asTTle.

Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O., & Sainsbury, M. (2001). *Framework and Specifications for PIRLS Assessment 2001*. Chestnut Hill, MA: Boston College, PIRLS International Study Center.

Glasswell, K., Parr, J., & Aikman, A. (2001). *Development of the asTTle Writing Assessment Scoring Rubrics*. Technical Report 6. University of Auckland, Project asTTle.

Hattie, J., & Purdie, N. (1998). The SOLO model: Addressing fundamental measurement issues. In B. Dart & G. Boulton-Lewis (Eds.), *Teaching and Learning in Higher Education.* (pp. 145–176). Melbourne, Aus.: ACER.

Kirsch, I., Mendelovits, J., & McQueen, J. (2000, July). *Assessing reading literacy in an international context: The OECD programme for international student assessment (PISA)*. Paper presented to the International Reading Association 18th World Congress, Auckland, NZ.

Limbrick, L., Keenan, J., & Girven, A. (2000). *Mapping the English Curriculum*. Technical Report 4. University of Auckland, Project asTTle.

Ministry of Education. (1993). *New Zealand Curriculum Framework.* Wellington, NZ: Learning Media.

Ministry of Education. (1994). *English in the New Zealand Curriculum.* Wellington, NZ: Learning Media.

Ministry of Education. (1996). *Exploring language: A handbook for teachers.* Wellington, NZ: Learning Media.

NCES. (2001). *The NAEP Reading Achievement Levels*, [World Wide Web]. Available: http://nces.ed.gov/nationsreportcard/reading/achieve.asp [2002, 23 January].