



Assessing the Readability of Māori Language Texts for Classroom Use

A thesis submitted in partial fulfilment of the
requirements for the Degree of
Master of Arts in Māori in the
University of Canterbury

by

Christine Mary Brown

School of Māori and Indigenous Studies
University of Canterbury

2009



He Mihi

Nei rā te mihi mutunga kore ki a koutou katoa i tautoko mai i tēnei mahi rangahau.

Ki aku kaiārahi i te rangahau, aku kaipenapena, ko Jeanette King, kōrua ko Alison Gilmore, e whakamihi atu ana tēnei ki a kōrua. Nā kōrua i ārahi mai ahakoa ngā pikinga me ngā hekenga o te mahi me ngā āhuatanga i puta mai. He pātaka mātauranga, he puna mōhiohio, kōrua.

Ki ngā kaikōkiri o te kaupapa nei, ki a Richard Benton kōrua ko Warwick Elley, nā kōrua hoki i akiaki mai, i tautoko mai. Nā kōrua i para te huarahi māku hei whai. Anō hoki ki a Mary Boyce, ko Paul Nation, ko Peter Keegan, he mihi whakawhētai ki a koutou katoa i āwhina i ahau.

E aku manukura o te hau kāinga, ko Ngāti Moki, ko Ngāi Te Ruahikihiki ki Taumutu, nā koutou ahau i korowai i taku mahi i ngā kura o tēnei takiwā. Kei te mihi, kei te mihi, kei te mihi. E moe te mata hī aua, e ara te mata hī tuna.

E te whānau manawanui o te kura, arā ko ngā kaimahi, ngā mātua, ngā tauira, te tumuaki me te pōari, nā koutou ahau i āwhina, i poipoi i te roanga o te mahi, ahakoa te aha. Mā koutou tēnei mahi, ko te tūmanako, he paku āwhina tēnei i a koutou. Ko koutou ngā poutokomanawa o tēnei whakatipuranga, hei oranga mō te reo rangatira o Aotearoa.

Abstract

This project sought to find a rigorous and manageable method for measuring the difficulty of texts in te reo Maori written for children, beyond junior reading material in Maori-medium educational settings.

The project examined a range of readability measures based on semantic and/or syntactic features of text, following the work of Warwick Elley (1969) and Richard Benton et al. (1995). Features such as the difficulty of content words, average sentence length, standardised type:token ratios and the use of function words were used in different combinations to create seven methods to measure text difficulty.

Teachers' and students' ratings of text difficulty, and students' scores on reading comprehension tasks related to the texts were used as criteria to examine the validity of the readability methods. The findings revealed that indices of either vocabulary load or lexical density when used in combination with the number of function types in the text, produce statistical significance with the criterion measures. Further research is needed to confirm their validity for use in Māori-medium classroom settings.

The Māori word lists developed for this project as the basis of the readability approaches have the potential for more widespread analyses of language proficiency measures for students in Māori-medium settings.

Table of Contents

He Mihi	i
Abstract	ii
Chapter One: Assessing the Readability of Māori Language texts for Classroom Use	6
Introduction.....	6
The Project Aims.....	7
Cautionary Note.....	9
Chapter Two: Estimating Readability	11
2.1 Introduction	11
2.2 Background	11
2.3 Word class and readability.....	13
2.4 Word class complexity.....	14
2.5 Scaffolding to develop reading frameworks.....	16
2.6 Critical transfer of literacy tools.....	17
2.7 Wider considerations for readability measures.....	18
2.8 Reading material published for Māori-medium education	19
2.8.1 <i>Ngā Kete Kōrero Framework Policy Project</i>	19
2.9 Primary Sources in the Literature	21
2.10 Summary	22
Chapter Three: Phase One, Word class difficulty, comprehension and vocabulary knowledge	23
3.1 Introduction	23
3.2 Background	23
3.3 Content words and comprehension.....	24
3.4 Replica Study	25

3.4.1 Participants	26
3.4.2 Method	26
3.4.3 Materials	27
3.4.4 Comprehension and vocabulary test.....	28
3.4.5 Results	28
3.4.6 Decoding accuracy, vocabulary knowledge and comprehension	37
3.5 Other findings which emerged from the study.....	42
3.5.1 Sentence length.....	42
3.5.2 Illustrations.....	44
3.5.3 Density of text and syntactic complexity	44
3.5.4 Reader interest and prior knowledge	45
3.6 Summary	46
 Chapter Four: Phase Two, Construction of the word lists	 47
4.1 Introduction	47
4.2 Methods of compiling word lists	47
4.3 Corpus materials.....	49
4.3.1 <i>A Corpus of Modern Spoken Māori</i> , Boyce (2006).....	50
4.3.2 <i>(Huia) Corpus of Māori Texts for Children</i> , (Huia Publishers, ongoing)	50
4.3.3 <i>Ko Ngā Kupu Pū Noa o Te Reo Māori, The First Basic Māori Word List</i> . Benton, Tumoana and Robb (1982), and <i>Basic Māori word list: levels six to ten</i> . Benton (1983)	51
4.3.4 <i>Te Kura ki Uta, Opotiki College Bilingual Education Programme Assessment and Evaluation Project</i> , Benton & Maxwell (1995).....	52
4.4. Method of prioritising data.....	53
4.4.2 Constructing the word lists.....	55
4.4.3 Grouping the words into score sets.....	56
4.4.4 Changes in usage	57
4.4.5 Months of the year, days of the week and numbers 1-10	58
4.4.6 Inclusion of word lists from <i>Te Kura Ki Uta</i>	59
4.4.7 Assigning values according to word frequency.....	60
4.4.8 Word sense.....	61
4.4.9 Word families.....	62
4.4.10 Plural forms.....	66
4.4.11 Word utility	66
4.4.12 Function word lists.....	67
4.5 Developing the Baseword Lists for RANGE	68
4.6 Word list totals	68
4.6.1 Dialectal and spelling variance	70
Summary	71
 Chapter Five: Phase Three, Linguistic analysis, ranking and selection of text series using word frequency.....	 73
5.1 Introduction	73

5.2 Using the software program <i>RANGE</i>	73
5.2.1 <i>RANGE</i> output data.....	74
5.2.2 Analysing text coverage using <i>RANGE</i>	75
5.3 Vocabulary Index: Method M1	77
5.3.1 Calculating a vocabulary load index.....	78
5.3.2 Inclusion of proper names.....	79
5.4 Vocabulary Index multiplied by function types: Method M2	81
5.5 Ranking previously levelled texts using <i>RANGE</i>	82
5.6 Ranking more advanced texts using <i>RANGE</i>	84
5.7 Selection of the text series for this study	84
5.7.1 Blue series	84
5.7.2 Orange series	85
5.8 Summary	86
 Chapter Six: Additional methods for estimating text difficulty	 88
6.1 Vocabulary Index: Method M1	88
6.2 Vocabulary Index x Number of function types: Method M2	89
6.3 Number of function word types: Method M3	90
6.4 Average sentence length: Method M4	91
6.5 Mean segmental type:token ratio: Method M5	93
6.6 Vocabulary Index + average sentence length: Method M6	95
6.7 MSTTR + number of function types: Method M7	96
6.8 Intercorrelation of methods	97
6.8 Summary	99
 Chapter Seven: Phase Four, Establishing Criterion Measures of Text Difficulty	 101
7.1 Teacher and student opinion	101
7.2 Participants	102
7.3 Ranking task	102
7.4 Student performance tasks	104
7.4.1 Constructing the student performance tasks	105
7.4.2 Administering the student performance task.....	106

7.5 Criterion measures	107
7.6 Intercorrelation of criterion measures	107
7.8 Summary	109
 Chapter Eight: Phase Five, Validation of methods to estimate text difficulty	 110
8.1 Introduction	110
8.3 Method of statistical analysis	110
8.4 The vocabulary index method (M1).....	111
8.5 Vocabulary index multiplied by function types (M2)	114
8.6 Number of function word types (M3)	116
8.7 Average sentence length (M4)	118
8.8 Mean segmental type/token ratio (M5).....	120
8.9 Vocabulary Index + average sentence length (M6).....	122
8.10 MSTTR + function types.....	124
8.11 Overall validity of methods	125
8.11 Summary	127
 Chapter Nine: Teachers' evaluations of the texts	 128
9.1 Introduction	128
9.2 Grammatical structures	128
9.3 Polysemy.....	129
9.4 Unknown words	129
9.5 Level of understanding and ability to re-tell	130
9.6 Use of idiom.....	130
9.7 Impact of dialect and borrowing.....	130
9.8 Reading strategies.....	131
9.9 Other comments.....	131
9.10 Summary	133

Chapter Ten: Conclusions.....	134
Emergent issues:	139
Corpora	139
Word lists	141
The conflicting nature of the text series.....	143
Participant numbers and range of texts.....	144
Tagging software	144
Future directions	145
Corpus development	145
Further testing.....	146
In Conclusion.....	147
References:.....	148
Appendices.....	159

Chapter One:

Assessing the Readability of Māori Language texts for Classroom Use

This thesis originated from my experience as a Resource Teacher of Māori working in Māori-medium education settings. It became apparent during my work in supporting teachers to develop literacy programmes, that there was a serious lack of structure guiding teachers in the selection of reading material for young developing readers. In discussion with colleagues the issue arose about finding a method of gaining some reliable insight into the difficulty of texts written for children, beyond junior reading material. This became the focus of this research project.

The importance of this type of study cannot be underestimated for young readers who are in transition from being supported readers to becoming independent and fluent readers. The turning point when students move from learning to read, to reading to learn is a critical time in the development of confident life long readers. The importance of careful guidance to support scaffolding is of paramount concern because students who struggle to read independently and confidently with good comprehension, will struggle to access all areas of the curriculum.

Introduction

The development of literacy in Māori-medium education reflects the premise that literacy in all forms is an essential element for using the Māori language in the modern world. This has become widely accepted and is now embodied in Māori-medium education and is supported by *Te Reo Matatini, Māori-medium Literacy Strategy* (2007). Research across all areas of Māori-medium literacy are in need of development, including those which can have direct application to the selection of materials for use in classrooms.

Levelling of texts was formally offered by the New Zealand Ministry of Education to the Māori-medium sector with the release of the teacher handbooks *He Purapura Handbook* (1995) followed by the *Ngā Kete Kōrero Framework Teacher Handbook* (1996, 1999 revised edition). A more detailed background to this levelling framework is provided in section 2.4. of this study. While there has been development of materials to support emergent and early readers in Māori-medium settings, there has been no formal development of the framework beyond these levels, nor is there any research guiding the selection or production of texts for newly independent and fluent readers. Furthermore, Te Pou Taki Kōrero (2007: 8) says there has been no reading material produced specifically for the fluency stage of the *Ngā Kete Kōrero Framework* during 2002-2005. Despite early intentions in the development of the framework, there has been no research completed to guide the development or levelling of material beyond the junior level. Therefore the current situation with regard to the selection of reading material beyond junior texts, relies entirely upon teacher judgement. If we compare how this stage of literacy development is treated in English-medium settings, we see that all of the reading material available for middle and senior primary school students is carefully organised. For example, a teacher in an English-medium school can electronically search all Ministry published *School Journal* material by year level, reading age, genre, author, and topic. The need for attention to this area of Māori-medium literacy development is urgent.

The Project Aims

This project aims to find a manageable way of objectively estimating the readability of texts in Māori for students who are becoming fluent readers and beyond. The most widely applied research in New Zealand guiding this investigation is Warwick Elley's (1969) 'noun frequency count method' for

estimating text difficulty in English. A range of methods derived from Elley's work is examined to determine the most appropriate for Māori-medium texts.

In order to achieve its aims, this study has been organised into five phases, each phase being dependent upon the outcomes of the previous phase.

The aim of Phase One was to briefly explore whether the word class of nouns can be isolated as carrying the weight of meaning in Māori texts. This is accepted as the basis for Elley's noun count method in English. In order to do this, a small study was undertaken based on work done in English by Marie Clay (1966) in which she analysed the word class of errors and self-corrections. This phase indicated that it would be more prudent at this stage to include all content class words into a modified Elley method for Māori texts, rather than use only nouns. This phase is detailed in Chapter 3.

The aim of Phase Two was to construct a set of word lists based on word frequency data to use for scoring the vocabulary load of a text. Designing a model based on Elley's method required a set of graded word lists. The findings of Phase One led to Elley's (1969) noun frequency count model being modified to include the complete group of open class words in the lists. The construction of the word lists was undertaken by amalgamating data from Māori corpus material available within the field of Māori education. The word lists increase in difficulty as determined by each word's frequency and range of occurrence in children's texts. The corpora which contributed to the construction of these word lists were those of Benton (1982, 1983), Boyce (2006), Huia Publishers (as at 2007), and Maxwell and Benton (1995). The development of the word lists is described in Chapter 4 and the lists themselves with their data are in Appendix 2.

Phase Three used the lists developed in Phase Two to select and rank texts. A modification of the Elley method was the core approach used to select two series of six texts from texts produced by Learning Media for middle school students.

The selection of the texts and ranking procedures are described in Chapter 5. At this point in the thesis it was decided to widen the approach and test other methods for estimating text difficulty. An additional five methods were explored and these methods are detailed in Chapter 6.

Phase Four aimed to establish criterion measures of reader opinion and student performance with which to validate the computed rankings of texts produced in Phase Three. In order to do this, the two series of selected texts were ranked for difficulty by a group of 10 Māori-medium students and 15 Māori-medium teachers. In addition, students undertook performance tasks and teachers completed questionnaires. The procedures used for collecting reader opinion and student performance are detailed in Chapter 7.

Phase Five was the final part of the study which aimed to validate the computed rankings using the six different methods, with the criterion measures of reader opinion and student performance rankings from Phase Four. The results of this validation process are presented in Chapter 8.

While this study has taken cumulative steps through phases, each with their own aim, the core aim of the project has been to establish the most valid ways of determining the likely difficulty of texts written in Māori for middle to senior school students. Chapter 9 summarises and concludes the study.

Cautionary Note

The work of transferring models of analysis between languages needs to maintain a critical approach. Nation and Worthington (1996) have found that special preparation of texts for second language learners of English bring about many factors needing special consideration and that it becomes difficult to

construct realistic texts. This project does not aim to encourage the deconstruction of the natural language that authors of children's texts in the Māori language may use. For the Māori language it is especially important that children are exposed to the natural language models that intergenerational transfer has to offer. It is important to clarify at this point that readability formulae need to be understood as only one component that contributes to wider text levelling procedures and it is also important to be clear that assigning reading ages to students in Māori-medium settings is not proposed in this study.

Chapter Two: Estimating Readability

2.1 Introduction

The aim of this chapter is to lay the foundation from the literature for the investigation into the way that content or message carrying words and their frequency of occurrence contributes to readability. Wider issues of readability are presented and the primary sources in the literature from the field of Māori language research are introduced. The maintenance of a critical approach in the transfer of literacy tools is also profiled, and the way that this study proposes to generate new information from previous work from within New Zealand is explained.

2.2 Background

Readability is usually defined as the judgement of how easy a text is to understand. While there are many criteria used in the measurement of readability, it is generally agreed that vocabulary knowledge plays a vital role in the comprehension of text (Anderson & Freebody 1981; Davis 1994; Elley & Croft 1989; Laufer 1997; Nation & Worthington 1996). There are two major components in written language. One is the set of words carrying the information (content words) and the other is the set of words that glue the information together and organise the grammatical relationships (function words). While the components of language are a closely knit code, content words stand distinct from function words in context. Words with lexical content are those that are most commonly classed as nouns, verbs, adverbs and adjectives. Function words, whose role in language is to express grammatical relationships, are the words such as pronouns, prepositions, particles, possessives, and conjunctions. Syntax, or the way a piece of writing is built, is acknowledged as a contributor to

levels of readability, and while a parallel consideration will be given to this aspect, Stahl (1986) says that the number of difficult content words in a text is still the strongest predictor of a text's overall difficulty. Difficult words are those that students have had little experience with and therefore they are often termed low frequency words. Thomas and Robinson (1977) identified these as 'stopper' words; words that cause the students to stop reading or lose meaning because the words are unknown. In support of this, Underwood and Schulz (1960) say that the higher the meaningfulness of a verbal unit, the more frequently that word has been experienced. Elley elaborates further in saying that:

The comprehension difficulty of a passage read will be strongly influenced by the frequency of occurrence in the English language of the key words in the passage, [that is] those words which carry the weight of meaning. (1969: 414).

For developing readers in any language, maintaining high levels of comprehension while developing skills for decoding text is critical. As students move into stages of fluency in reading, the risk of gaps appearing between their ability to decode the text and their ability to unlock meaning from it are heightened. This is especially true for developing readers of Māori language. Because of its phonemic regularity, Māori is a comparatively friendly language for beginner readers. However, quite quickly the reader's ability to 'say' the words on the page outstrips their ability to talk about what they have read. Therefore, the vocabulary load and comprehensibility of instructional texts needs to be monitored carefully, to ensure strong scaffolding is provided for reading development. This is important for reader confidence and will optimize regular success for the reader. Reading for meaning is paramount and is interdependent upon teacher knowledge about text selection. Laufer (1989) says that no more than 5% of words in a text should be unknown to the reader in order to read with adequate understanding.

2.3 Word class and readability

Among others, the work of two New Zealanders has contributed much to the field of understanding issues of readability. Clay (1966) found in her analysis of errors made by developing readers in English, that nouns were the most difficult word class to predict in a flow of text and also the most difficult to self-correct once an error is made. Elley (1969) concurs with this, describing nouns as being the word class surrounded by less redundancy than other parts of speech, making their comprehension more critical to understanding the text. Therefore, Elley concludes, the key words that carry meaning in English texts are content words and more specifically, are nouns. Phase One of this study was implemented to get an indication of whether this finding is also true for Māori language texts.

Elley's noun frequency count method is a significant tool which contributes to the levelling of educational materials produced for English-medium settings. Warwick Elley (private communication) has encouraged the exploration of this method for the Māori language, in the hope that the benefits of this approach can also be employed by the Māori-medium sector. Nevertheless, Elley warns, that objective measures of readability which use only empirical approaches to measuring the suitability of texts have limitations. While they make a worthwhile contribution to the placement and selection of texts to be used in literacy programmes, they are not intended as stand alone mechanisms. Furthermore, there are certain text types that are not recommended for application of the noun count. These include poetry and non-fiction texts with technical language.

It is recognised that there are many features of texts that need to be considered in text selection. However, Pitcher & Fang (2007) state that the most rudimentary estimate of text difficulty is knowing the vocabulary burden that a text contains. There are also many ways that text complexity can be measured. Readability formulae were first developed in the 1920s in the United States, while the

broader spectrum approach incorporating a range of text support factors known as levelling, began even earlier (Pitcher & Fang, 2007; Fry, 2002). In New Zealand the work of Elley has had longstanding prominence in the quantitative aspects of the levelling of children's reading material. This method of estimating the difficulty of reading material and its associated reading age ratings, is known as the 'noun frequency method' and has been the mainstay of levelling the School Journal series and the national standardised Progressive Achievement Tests (*PAT*) for reading comprehension. Notwithstanding the validity of this method, Elley and Croft (1989) draw our attention to the limitations of objective readability measures and reinforce the wide lens approach advocated by Clay (1991). Two methods tested in this study contain some basic modifications to the noun frequency method. These modifications are mainly due to insufficient evidence that the word class of nouns is the key burden factor for the Māori language, and a lack of software to make a noun count analysis manageable.

2.4 Word class complexity

Languages are very complex systems, which are difficult to explain within a rigid set of rules. Native speakers of English will be instantly familiar with the rule breaking complications that the English language presents. The Māori language is no less complex, and while it is possible to generally separate out the behaviour of content words from the smaller but more frequent group of function words, it is more problematic to then classify the lexical bases into word classes. The following examples from scholars in the field illustrate this point.

Williams (1992: xxxiii) notes that while there are examples of a shared word class in English, most words in Māori may be used in more than one word class. An example of a shared word class in English is the noun 'book' which can mean 'a book' that can be read but can also take on the verb class of 'book' as in to 'book a seat'.

Biggs (1998: 54, 55) describes features of the words he specifically categorises as nouns for Māori. Furthermore, he describes a further class as ‘universals’ being the largest class of Māori words because they can be used in both verbal and nominal phrases. In Māori, the word *waiata* can be used as a noun (song), as a verb (sing), and as a modifier (singing group). Boyce (2006: 31) uses the following examples to illustrate this point from the *Māori Broadcast Corpus* hereafter referred to as the *MBC*:

As a verb:

Nā, ka noho ka waiata.

As a noun:

He maha ētahi atu waiata mai i te rōpū Wai-hīrere.

As a modifier:

I whakatūria tēnei rōpū waiata, ā, nuku atu i te whitu tekau ngā tau ināiane.

Biggs does not allocate a separate class for verbs. The class of statives (closely aligned with adjectives), locatives, and personals are further categories of his which all constitute lexical bases or content words.

Bauer (2003: 65) states that “the analysis of Māori vocabulary into parts of speech or word classes is an area where there is disagreement in the scholarly community”. She disagrees with some of Biggs criteria for word class groupings and presents another set of descriptors which does include a class for verbs.

Benton et al. (1982: 5) describe content words as those “expressing some independently definable object, action, state, belief etc”. Boyce (2006: 267) identifies the content words as those that operate “as bases – nouns, verbs, and adjectives or modifiers”.

For the purposes of this research, the general class of Māori content words that are agreed by linguists as being the lexical bases including nouns, verbs, universals and modifiers are to be the focus of the word frequency counts. Richards (1974) also supports the idea that modifiers or words which help define other words and affixed forms are important in carrying message.

2.5 Scaffolding to develop reading frameworks.

Theories of scaffolding in the development of print literacy are agreed upon as an important element in Māori-medium programmes. This is supported by Hohepa, Smith, Smith & McNaughton (1992), Pere (1991), Tangaere (1997), Skerrett-White (1995), and also in the field of wider bilingual education, Cummins (cited in May, Hill & Tiakiwai 2004), and Ellis (2005). Regardless of how the developmental process is described, by lines, charts, poutama, spirals, or rubrics, the fact remains that learning is a progressive and cumulative process across all cultures.

The scaffolding of print material in education is variously described as levelling, grading and indexing. Giving teachers indicators of where texts lie in relation to others, enables them to support and guide students toward new skills and also signposts ground that can be covered again for consolidation. Experienced teachers of reading understand that the journey sideways is a critical part of strengthening and facilitating the journey forward. Access to levelled texts means that teachers can increase their sense of direction, know which way is forward, and which way is sideways. Within this understanding, levelling is not feared as a mechanism to accelerate children in a high pressured or linear way. Children travel different pathways and at different rates toward becoming literate. Both shared and instructional reading using levelled material can be viewed as working within what Vygotsky (1978) described as zones of proximal

development. By providing interaction and guiding a reading task that the student would otherwise be unable to achieve, a text can enrich the language experience of the student and model for them further complexities in both structure and vocabulary. Krashen's (1989) input hypothesis theory emphasises that the more comprehensible aural and written input is, the greater the language acquisition. Ensuring text input is comprehensible is an important part of scaffolding. Clay (2007) says that 95 – 100% accuracy will allow an easy reading of a text while 90-94% accuracy is deemed to provide challenge sufficient to allow learning from errors to occur. Below 89% means the reading experience will be hard. Nation (2004) recommends there be less than one unknown word in every 20 words read for reasonable comprehension to be achieved from a text. It follows then that it will be useful to know which words, and how many of them are contained in a text that are likely to be unknown, especially in the case of second language learners. Fry (2002: 291) says that:

Selecting books by a readability formula or leveling procedure is only one of the many teaching techniques that a reading teacher should use, but it is one that will help many students have a successful learning experience.

2.6 Critical transfer of literacy tools

Longstanding frameworks for assessing the readability of instructional material used in the teaching of reading are embedded within practice for English-medium programmes (see for example the assigned levels in the *New Zealand School Journal* series).

It is understood that neither international nor local linguistic research findings can automatically be applied from one language to another. For reasons of cultural legitimacy and linguistic uniqueness, there is always an extensive and critical

process of testing, modification and adaptation of some aspects, and in some cases complete rejection. Many tools used in the teaching of reading which monitor the reading behaviour and competence of New Zealand children are now considered crucial components of any quality programme and some of these tools have been critically reconstructed and transferred to assist literacy development in Māori-medium settings. For example, Dame Marie Clay's extensive work on diagnostic assessment which includes the Diagnostic Survey known in English as the Six Year Net has been adapted for Māori-medium into *He Mātai Mātātupu* (Rau, 1998); School Entry Assessment has been transformed into *Aromatawai Urunga-ā-Kura* (Ministry of Education, 1999); running records have been adapted into *Pūkete Pānui Haere* (Rau, 1998); and three minute running records into *Iti Rearea* (Glynn, Harawira, Durning, 1993). These tools are now extensively used in the Māori-medium sector to monitor reading behaviour and competence. While approaches to developing literacy for Māori-medium are emerging from the transfer of proven methods for first language speakers of English, there are strong views that models which exist for English-medium literacy cannot simply be transferred uncritically into Māori-medium literacy. To do so would, as Rau (2001: 2) clearly states, "seriously compromise the integrity, the reliability and the validity of such measures and assume pedagogical and cultural compatibility".

2.7 Wider considerations for readability measures

It is recognised that the readability of any given text is the result of a combination of many factors. Surface features such as the typeface, density of layout, design, strength of illustrative support and sentence length, are all contributors. The reader's personal and conceptual interest level, background knowledge of the content, the experience of the genre, the syntactic complexity, and the number of key words used which are unknown to the reader, are also important factors for consideration. The *Ngā Kete Kōrero Project* (1995), and *The Levelling of PM*

books and Benchmark kits, (2003) both strongly support the inclusion of the abovementioned factors.

2.8 Reading material published for Māori-medium education

From as early as 1960, with the publication of the Māori language readers, *Te Wharekura*, through to the present day, written material has been published for middle to senior school students in Māori. There is now a large collection of material which spans wide-ranging levels of difficulty, sometimes within the same publication (for example, *He Kohikohinga*, *Ngā Kōrero*, *Te Tautoko* and *Te Wharekura* series). For this material to be utilised with professional confidence concerning the level of challenge it will present to a student, each text in these series should be given some kind of readability rating. All students and teachers in New Zealand schools working with material published in English can access this type of guidance.

2.8.1 *Ngā Kete Kōrero Framework Policy Project*

Te Puni Kōkiri (Ministry of Māori Development) published the *Ngā Kete Kōrero Policy Project* (Benton, Glynn, Kapa, Murphy, Berryman, Hindle, Rau C, 1995). The project was established to address a previously identified issue facing Māori-medium education; that of the urgent need to provide a levelled series of reading texts published in Māori. The project design incorporated the development and trialling of a framework appropriate for classifying existing and new texts published in Māori, the production of new resources, and finding sponsorship to support ongoing production. This is the only formal framework that has been developed for levelling reading material in Māori-medium education. Unlike the grading of levelled reading material in English, there is no chronological age or 'grade level' assigned to Māori-medium materials. Bishop, Berryman, and Richardson, (2001: 34) make reasons for this clear:

Students entering Māori-medium schools, do so from a far broader language continuum than do the majority of their English-speaking peers who are entering English-medium education. It is also likely to be true that most students entering Māori-medium schooling are encountering Māori as a second rather than a first language. This needs to be taken into account when making any assessment of Māori literacy and in the devising of teaching and learning strategies and resource materials.

Failure to take into account the stage of the language competence of the learner, rather than their age, seems likely to lead to students appearing to perform below expectation. In fact, such students may well be performing at, near, or above where they ought to be given their language exposure, and this may bear scant relationship to their chronological age.

The main aim of the framework team in this project was to “develop a classification system for basal reading texts” (Benton et al., 1995: 17). The framework team acknowledged that they were prevented from exploring Elley’s frequency count method at the time of the project because of the absence of a Māori language corpus large enough to provide reliable data on word frequency. Benton et al. found that trying to establish a measure of readability based on vocabulary level and word type did not produce useful results. However, a more complex formula using word lists drawn from *Ko Ngā Kupu Pū Noa* (Benton et al. 1982, 1983) and the *National Curriculum Draft Statement for Māori Language* (Ministry of Education, 1994), was more promising. The elements in the formula included: basic vocabulary, other frequent vocabulary, totals of word types to tokens, the number of particles, the number of different pronouns, demonstratives, and relative location or identity. These all contributed to scoring on three indices involving richness, syntactic complexity, and relational

complexity. Unusual vocabulary was then factored into the final combined score. A computer programme was designed to calculate this formula. The formula used by Benton et al. as described above, is indeed complex and very thorough in its linguistic analysis. This work appears to be the only work to date which attempts to generate objective and computational data on the likely difficulty of texts in Māori. Using their method of analysis, the framework team identified eleven levels which were then grouped into five main levels, some of which contain sub-levels. The levels were named Harakeke e, Harakeke i; Kiekie a, Kiekie e, Kiekie i; Pīngao a, Pīngao e, Pīngao i, Pīngao o; Miro; and Whatu. Benchmark books were selected which provided a yardstick for other material to be levelled against for placement onto the framework. Currently, the framework only applies to reading material produced for junior classes, up to a stage described in the framework as a fluency level named 'Miro' The following level, 'Whatu', had no benchmark text assigned.

Before the development of the noun frequency count method, Elley and Croft describe the task of estimating the difficulty of reading materials as having been:

Entirely unsatisfactory; most are cumbersome to apply time consuming, arduous and the results were by no means consistent. (1989: 7).

The noun frequency count method was proven to simplify the process without compromising effectiveness. It has been 25 years since Benton et al. generated their early corpus work in this area. With the new corpus information that is available, it is timely that word frequency data are revisited.

2.9 Primary Sources in the Literature

The most significant reference points in the research and literature and the main sources which guided this project are discussed separately as they underpin each phase. They include: Benton, Glynn, Kapa, Murphy, Berryman, Hindle, &

Rau, (1995) (in section 2.7.1); Elley (1969), and Elley & Croft (1989) (in section 4.1); Boyce (2006) (in section 4.3.1); Heatley, Nation & Coxhead (2002) (in section 4.3.1); Huia Publishers (unpublished corpus) (in section 4.3.2); Benton, Tumoana & Robb (1982, 1983) (in section 4.3.3), and Benton & Maxwell (1995) (in section 4.3.4).

Other key guiding elements from the literature are those which focus on vocabulary content, and measures of comprehension of text. For example: Rye (1992), Nicholson (1991), Riley (1973), Pugh & Brooks (1986), and Clay (1966) all contribute a rich body of knowledge concerning the most reliable ways of measuring comprehension, and therefore the comprehensibility of texts.

2.10 Summary

While there is little other work which covers the exact scope of this study, there is a body of literature guiding the construction of word lists, and also a growing pool of corpus work now available which makes conducting this work possible for the Māori language. This means that robust word frequency data are now available for use in estimating text difficulty. This is explained further in Chapter Four. There are New Zealand-based initiatives in Warwick Elley's noun frequency count method, Heatley's et al. software tools and Richard Benton and his colleagues have provided a very solid basis from which this type of work can now be furthered. It can be seen from the literature that there is a need in Māori immersion settings to safeguard the reading experiences of developing and fluent readers alike. The literature signals the importance and benefits to be gained from attention being given to vocabulary development and there is a gap in the provision of professional knowledge for teachers about text difficulty. Therefore, this work proceeded within a firm base of research and established need, and was to be undertaken with an awareness of a critical approach to the transference of other models onto the Māori language.

Chapter Three: Phase One, Word class difficulty, comprehension and vocabulary knowledge

3.1 Introduction

In this chapter, a minor replication of Clay's (1966) error analysis of running records is presented. Elley (1969) conducted a study using cloze testing which showed that nouns had the lowest rate of correct replacement. In addition to this, Clay's (1966) work which analysed more than 10,000 errors made by beginning readers using the Ready to Read books, showed nouns to be the most problematic word class for children to self-correct. This encouraged Elley to test graduated noun frequency lists as a key tool in measuring readability in English. It was therefore decided to analyse a small set of running records for this project to see if a specific word class for Māori could be isolated as showing similar trends.

3.2 Background

The Elley (1969) noun frequency count method, identified clearly that the word class of nouns contributed significantly to the readability of a text. Elley & Croft (1989) investigated the usefulness of several potential measures of readability based on the frequency data collected from a large corpus of New Zealand children's writing. Five independent studies tested the following combinations: noun frequency; noun plus adjective frequency; noun plus verb frequency; verb frequency; adjective frequency; the Lorge formula (Lorge 1944); the Dale-Chall formula (Dale & Chall 1948); the proportion of abstract nouns; the proportion of unfamiliar words; sentence length; and ratio of prepositional phrases. Of all of the methods tested, the noun frequency level was the one that consistently correlated most highly with teacher and pupil opinion of the difficulty of English texts.

3.3 Content words and comprehension

The key message carrying words are widely recognised as those known as content words. The remaining set of words are known as function words, those which express grammatical relationships, and whose meaning is usually dependent upon the context. Function words have been shown by all corpus work on the Māori language to be the most frequent. See, for example, Benton et al. (1982), Boyce (2006), Harlow and Thornton (1986). Function words are, therefore, the class of words most well-known to children. In contrast, the content word class has a very wide scope for being unfamiliar or being low frequency words. They are sometimes called 'information-content' words, which accurately describe their role in carrying information. For example, imagine a set of English function words like this: there, is, a, the, up. This set of words contains no message at all. However, in contrast if you had a set of content words, you could glean some kind of message such as: cat, chasing, bird, tree.

A primary idea in considering the key message carrying words is the frequency aspect of those words. Elley argues that it is a reasonable speculation that meaningfulness (in the sense of ease of understanding) is also a correlate of frequency of exposure. Elley lists other studies which have confirmed this position such as; Bormuth (1966), Dale and Chall (1948), Forbes (1952), Gray and Leary (1935), Lorge (1948), and Spache (1968). In addition, Elley cites more recent studies by Harrison (1980) and Davison (1985) which all conclude that vocabulary is the most important factor in determining readability. All of the above have found high correlation coefficients between comprehension difficulty and the proportion of unfamiliar words in a text. This background about the importance of content words and their frequencies, and the bearing this has on comprehension is the evidence supporting the approach to this current study.

3.4 Replica Study

Marie Clay (1966) analysed self-corrections made in children's oral reading and found that the class of words most difficult to self-correct were nouns (only 21% self-corrected) followed by adjectives (33%) and verbs (35%). The function word classes scored higher rates of self-correction. Elley (1969) found this correlated with his analysis of cloze testing, that nouns proved to be the most difficult class of words to replace correctly. Adjectives and adverbs were others which showed low correct replacement results.

Elley (1969) provides a description of his method of assessing the readability of children's reading material using word frequency measures to rate the nouns in a given passage. According to Elley & Croft (1989) there is a strong correlation of meaning and understanding of verbal units with the frequency of exposure. It would, therefore, be fair to believe that high frequency words which children experience in their listening vocabulary are most likely to be those which they easily assign meaning to while reading. Measuring the frequency level of the nouns in a text was found to provide a highly valid estimate of text difficulty. Low frequency nouns were shown to be the strongest indicator of the level of challenge that a text would carry.

In order to test the 'noun frequency level' as being the key word class in readability for Māori, Clay's (1966) analysis of running record errors was considered to be the most manageable, worthwhile and appropriate groundwork for this study for two reasons. Firstly, the generation of running record data for the field site school where data were collected, provided teachers with fresh baseline information to begin their year. Within the framework of kaupapa Māori research it is essential that the key stakeholders in the research are also the primary beneficiaries, not only at the conclusion but, also where possible, during

the process. This was seen as one such opportunity to benefit the school while the research was in progress. Information was provided to the teachers on their students' reading behaviours and levels of comprehension which provided evidence to guide their teaching focus. Secondly, it helped maintain a critical approach to the application of the noun frequency method which has its roots in English, to the Māori language. It was considered important to explore whether the same word class isolation of nouns as posing the greatest reading challenge, would be true for Māori. This was to be judged by the criterion of self-corrections as Clay (1966) did.

3.4.1 Participants

This small replica of Clay's (1966) analysis involved a group of 35 year 5 – 8 students at a kura kaupapa Māori. The students of mixed gender ranged in age from 8.1 to 12.6 years. The reading levels of the students had previously been determined by their teachers as ranging between *Pīngao* and *Miro* levels of the *Ngā Kete Kōrero Framework* as explained in section 2.8.1. This produced three groups: 16 students at the *Kete Pīngao i* level, 8 students at the *Kete Pīngao o* level and, 11 students at the *Miro* level. The students were all engaged for the term in a whole school topic about the sea.

3.4.2 Method

The three groups of students were tested using three minute running records, an assessment tool used in Māori-medium education known as '*Iti Rearea*'. The texts were previously unseen so as to get an accurate measure of the students' ability to engage with the text unassisted, and observe how they negotiated decoding low frequency words. Students were then given a set of comprehension questions which were administered orally. The testing was digitally recorded for analysis later. No marking took place during the testing, in order to reduce the formality of the process and to allow the researcher to mark more accurately

later. To contribute toward reliability, the testing was all conducted in the early part of the day by the researcher. The researcher is known to all of the students, many of whom have read to the researcher previously.

3.4.3 Materials

The texts used were from collections produced for Ministry of Education by Learning Media and were selected for:

- i) Their relevance to the students' current topic of study;
- ii) Their level of difficulty as tentatively placed by Benton et al. in the *Ngā Kete Kōrero Policy Project* (1995); and
- iii) The presence of potential 'stopper' or low frequency content words that they contained.

The texts for the three groups were:

Kete Pīngao i: Waimarie He Moemoeā Noaiho, by Charles Nicholson (1994). In *He Kohikohinga 11*, p.10. (210 words). (Abbreviated as *KPi*: HK 11:10).

Kete Pīngao o: Te Aumoana, by Alison Robinson, translation by Roka Paora (1974). In *Ngā Kōrero 2*, p.26. (366 words). (Abbreviated as *KPo*: NK 2:26).

Kete *Miro*: Kaitiaki Kaimoana Interview of Eddie Waitoa, translated by Wiremu Kaa (1990). In *Ngā Kōrero 8* p.26. (partial text of 487 words). (Abbreviated as *Miro*:NK 8;26).

The texts are contained in Appendix 1.

3.4.4 Comprehension and vocabulary test

Students read the text aloud to the researcher and then listened to a recorded reading while visually following the text. The recording used strong intonation and pausing to emphasise meaning. Students then answered the comprehension questions orally. The decision to administer the comprehension test after a second exposure was based on discussions with the teachers, some who are first language speakers of Māori. A native speaking teacher explained that when they read Māori, the first exposure gives them a feel for the speaker and their background, the dialect, the depth of language used and the wairua or spiritual dimensions of the author. They explained that it is not until the second reading that full attention to content occurs. This experience is also consistent with comments from other teachers as second language learners. Most of the students involved in this study were second language learners of Māori. Swain's (1985) 'bias for best' approach was taken for the students and they were given a second exposure to the text with the recorded reading prior to testing comprehension. The comprehension test followed a set format of questions containing simple recall of detail through to inference, evaluative and creative response questions. There was a vocabulary probe in the test which contained a selection of the low frequency content words, and students orally explained their understanding of the words. Words were not tested in isolation but were discussed and identified in context and re-read where they had occurred in the text. The comprehension probes are contained in Appendix 1.

3.4.5 Results

The reading accuracy and behaviours were recorded using the conventions developed by Rau (2004) from Clay's Diagnostic Survey. The texts were analysed into word classes of nouns, verbs, "others" (which included modifiers

and onomatopoeiac words), and function words. These classifications were determined by how the word was being used in context. The three student groups and the texts they read are reported on separately.

3.4.5.1 Kete Pīngao i

The texts were analysed into word class groups in order to discover the composition of the texts so that the self-corrections from running record data could be interpreted accurately in relation to the distribution of word class throughout each text. Table 3.1. shows the word class analysis of the text *Waimarie He Moemoeā Noaiho* (Nicholson, 1994).

Table 3.1: The number and percent of words falling into each word class for the text: *Waimarie He Moemoeā Noaiho*

Word class	Number	Percent
Verbs	29	14
Nouns	40	19
Other content words	13	6
Function words	128	61
Total	210	100

It can be seen that word classes are not represented evenly throughout a text. The potential for error in this text was the greatest for function words (61%), followed by nouns (19%), verbs,(14%) and other content words (6%). To recognize the potential error that exists in each word class, it is necessary that self-corrections are also considered in proportion to the potential for miscues. The self-corrections made by the 16 students in the *KPi* group were analysed into the same word class groups as described above, and are presented in Figure 3.1 as percentages.

Figure 3.1: Proportion of self-corrections made to miscues by the *KPi* group for each word class

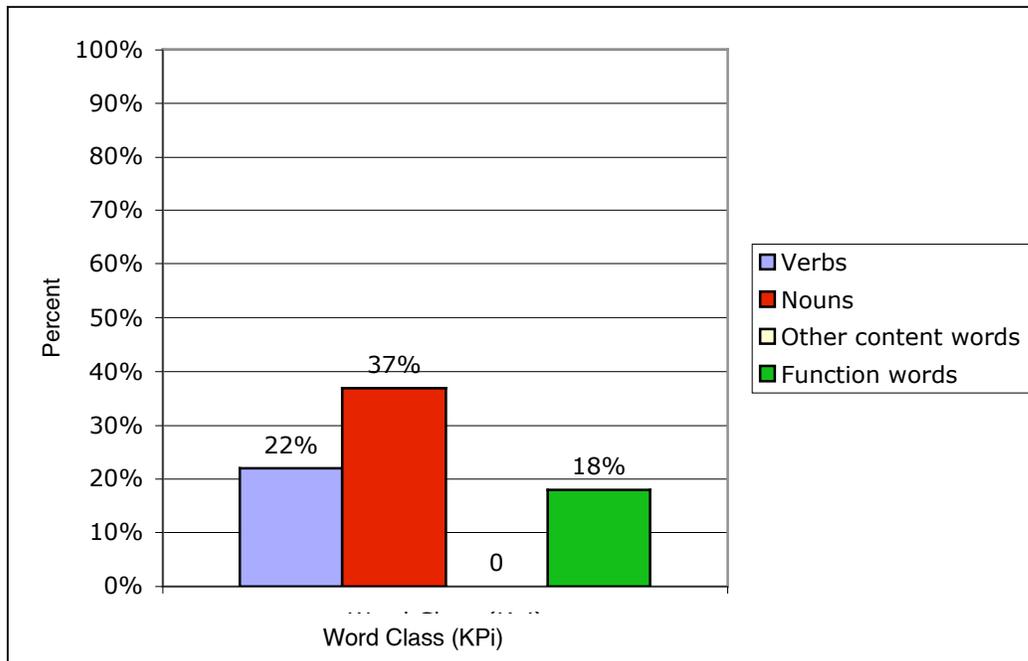


Figure 3.1 shows that nouns were the word class most successfully self-corrected (37%), followed by verbs (22%) and, in this instance there were no self-corrections (and no miscues either) for the group of words classified as “other”. The findings for this sample are in contrast to those of Clay (1966) who found that nouns were the most problematic for students to self-correct in English. Also in contrast to Clay’s (1996) findings, Figure 3.1 shows that in this sample, students made fewer self-corrections of function words (18%) than verbs or nouns. An explanation for this could be that as readers progress to fluency (i.e. performance free of undue pauses and false starts), they often become casual about words that don’t carry significant weight for understanding. In addition, Rye (1983) says that studies have shown that fluent readers make fewer fixations per line, and tend to focus on the longer groups of letters which are mainly content words. The small function words such as *i*, *ki*, *a*, *o*, accounted for most of the

function word errors and generally had little impact on reading flow and intonation reflecting understanding. In listening to the spoken language of students, it is also often the small function words that are missing in their spoken grammar. Analysis of their written work shows the same pattern. This could be a reflection of the linguistic stage of their development of grammar. Perhaps their spoken grammar patterns are not yet strong enough to make errors ‘sound’ incorrect as they read. Because readers did not stop to self-correct these types of error, there was a minimal impact on reading flow.

3.4.5.2 Kete Pīngao o

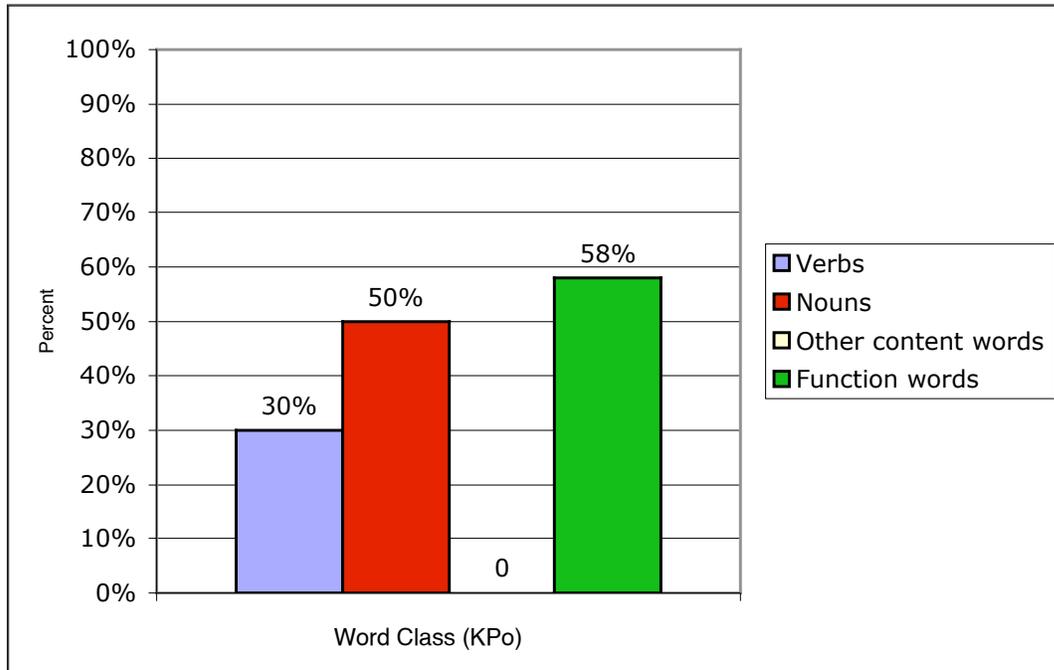
Table 3.2 shows the word class data for the text *Te Aumoana* (Robinson, 1974) which was the text used for the *KPo* group of 8 students. These data are also presented as proportions of the total text.

Table 3.2: The number and percent of words falling into each word class for the text: *Te Aumoana*

Word class	Number	Percent
Verbs	37	10
Nouns	92	25
Other content words	14	4
Function words	225	61
Total	368	100

This text shows similar proportions of distribution across the word classes as the previous text, with the function word class making up 61% of the text, followed by nouns, verbs and “others”. Figure 3.2 shows how well the students for this group self-corrected each word class, presented as proportions, to maintain the integrity of representation for each word class in the text.

Figure 3.2: Proportion of self-corrections made by *KPo* group as a percentage of the miscues for each word class



As for the previous groups, the results in this sample also showed that nouns were the most successfully self-corrected group of content words. These findings are also in contrast to Clay's (1966) findings for English. This group of students showed a good rate of self-correction for function words which probably reflects the growth in their knowledge of grammar. They were least successful at solving miscues for verbs.

3.4.5.3 Kete *Miro*

The final set of results are those from the *Miro* group of students who read the text *Kaitiaki Kaimoana* (Waitoa, 1990). The running record analysis, however, was not done using the whole text of 986 words but was undertaken on a reduced text. This was because using a timed running record, the students comfortably read only 487 words in the three minutes allowed. Table 3.3 shows

the word class analysis and Figure 3.3 show the data as proportions of the reduced text.

Table 3.3: The number and percent of words falling into each word class for the text: *Kaitiaki Kaimoana*

Word class	Number	Percent
Verbs	50	10
Nouns	115	24
Other content words	22	4
Function words	300	62
Total	487	100

This text also shows the same trend as the others, the pattern for function words is consistent with the findings of Boyce (2006) who says that high frequency function words make up a large proportion of any text, and that this is consistent for many languages.

The students reading the *Miro* level text made several whole line skips and re-runs of whole lines. These insertions and omissions were recorded on the running record schedule and included in the formula to calculate accuracy data, but they were not included in the word class miscue analysis. This was because insertions were not actual word types found in the text, and omissions of words occurring through the skipping of lines, showed that students did not engage in any way with decoding these words. Including these 'skipped' sets as miscues would have skewed the data away from observing actual 'stopper' word types. The prevalence of line skips from this group, however, did show the impact that increased density of text at higher levels can have on reading behaviours.

Figure 3.3: Proportion of self-corrections made by *Miro* students as a percentage of the miscues for each word class

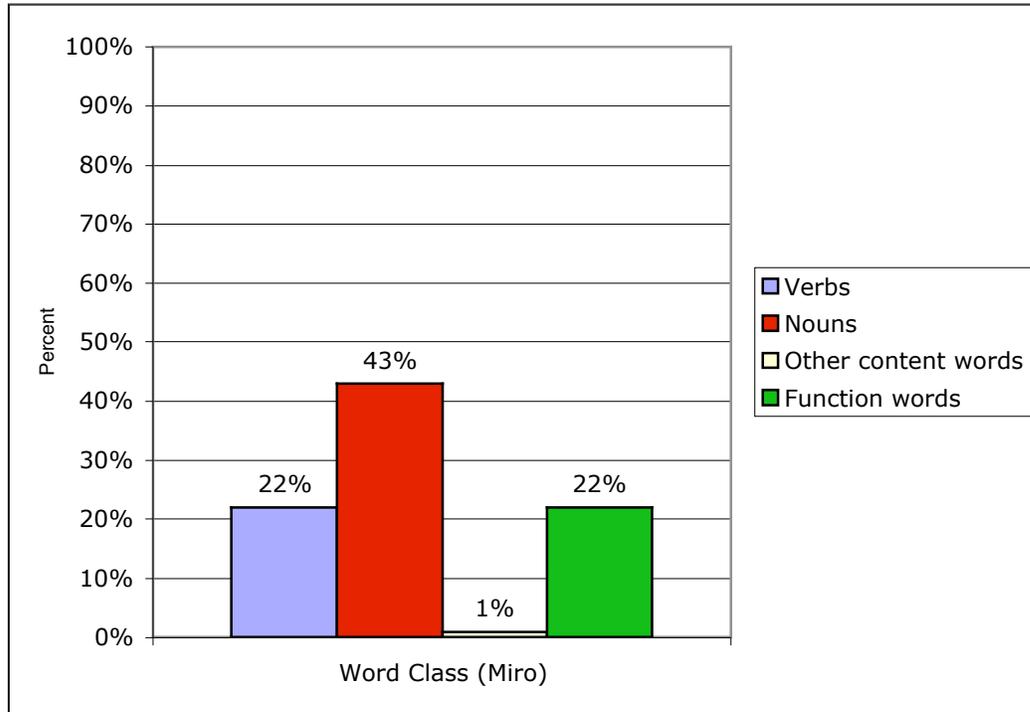


Figure 3.3 shows that for this group, nouns were not the most problematic class of words to self-correct, and that the verbs and function words were equally successfully self-corrected. The lower rate of self-correction for this group on the function word class might be explained by the presence of more complex sentences. However, it also leads us to pay more attention to the role that function words play in estimating text difficulty. In addition, they had almost twice the difficulty in solving miscued verbs than nouns.

3.4.5.4 Comparison of distribution and self-corrections by word class across Phase One texts

The three texts used for this section of the study showed a strong pattern of similarity with each other in the distribution of word classes as shown in Figure 3.4.

Figure 3.4: Proportions of word class across the three texts used in Phase One

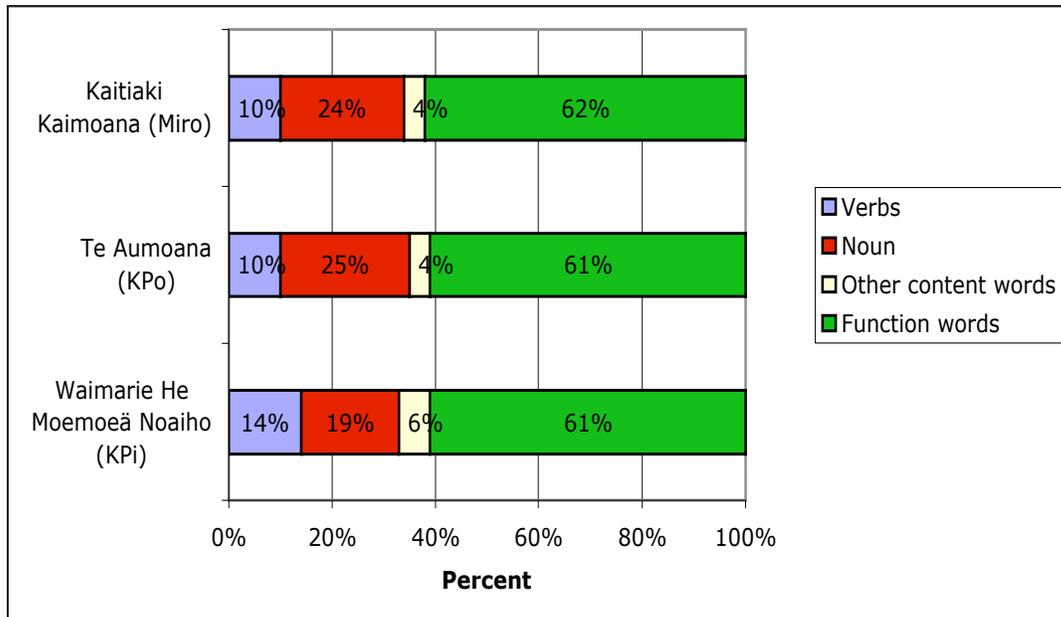
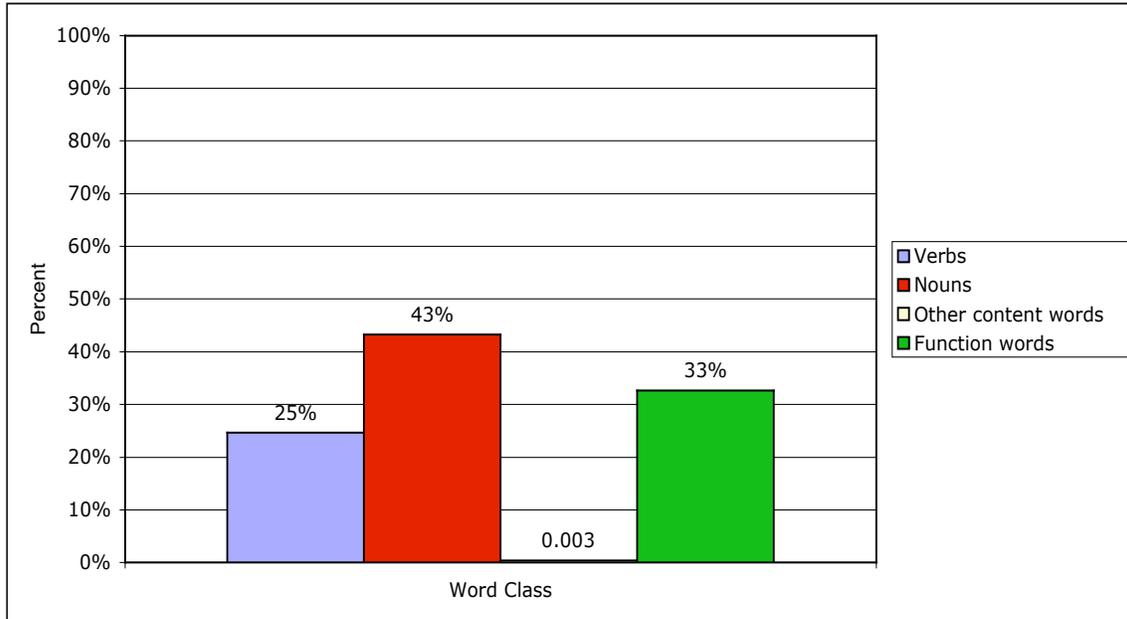


Figure 3.4 shows that function words make up just over 60% of each text, and also shows similarities in proportions of nouns (19-25%), verbs (10-14%) and “others” (4-6%). According to Hudson (2007) this distribution of word classes is consistent with most texts in any language. He states that two major English corpora show that common nouns make up 24% of the words in informational texts, similar to the *Miro* and *KPo* texts used here, and that imaginative texts like the *KPi* text, show a proportion of 19%. The patterns of distribution shown for the above texts show remarkable similarities to Hudson’s findings. The nature of word class distribution for texts written in Māori for children would be interesting to explore further, but because there is no software currently available for tagging text into word classes in Māori, this is not feasible beyond a few simple manual analyses at this point in time. Boyce (2006) states that function words make up the largest proportion of tokens in the *MBC* and other Māori corpora - the same pattern occurs for other languages.

Figure 3.5 shows an averaged proportion of self-corrections for each word class for the whole group.

Figure 3.5: Self-correction rates for word class as an averaged proportion for all three student groups during Phase One



This data clearly shows that nouns are not proving to be the most problematic word class for any of these students reading in Māori, and that verbs and function words play a greater role in generating reading miscues.

To summarise the above findings, although the word class of nouns is more strongly represented in the texts than verbs, and therefore present more opportunity for miscue and hence self-correction, the self-correction rates for nouns is still greater than the difference in representation. For example, in the *KPi text Waimarie He Moemoeā Noaiho*, nouns are represented 5% more than verbs. However, the self-correction rates for nouns is 12% more than verbs. This pattern is repeated across all three groups. While this replica study may be too small to be conclusive about these patterns, these findings were sufficient to

steer the research away from focusing purely on nouns in further phases of the study.

3.4.6 Decoding accuracy, vocabulary knowledge and comprehension

Hu and Nation in Nation (2001) concluded that in a written fiction text, learners needed to know 98% of the running words in order to gain adequate comprehension. This brings tight constraints to bear upon the texts that are presented to children to read that will ensure a good measure of understanding and success. It also supports the importance of having a measure of the density of low frequency vocabulary contained in a text. The basic idea of familiarity and vocabulary knowledge is that of the exposure theory, which explains the development of receptive or hearing vocabulary (Miller and Gildea, 1987; Stenner, Smith and Burdick, 1983). Educators are generally in agreement that productive language ability is basic to reading comprehension. Productive language ability is heavily dependent upon the vocabulary bank that a speaker has available to draw upon. A poorly stocked vocabulary bank impacts on understanding for both listening and reading. Krashen's (1989) comprehensible input hypothesis concurs in stating that when too much input vocabulary is unfamiliar, loss of meaning occurs. Nicholson (1997) describes this as a bottleneck problem, which, while it may not always affect overall comprehension, it does interrupt precise understanding at the sentence level. This difficulty was borne out during the comprehension testing. While some students got the overall gist of the story, they still came away with some key content information missing. There is a danger for students and teachers alike, in believing that gleaning the general meaning is an acceptable level of understanding. For second language learners this can often become the goal; that is, just to get the general idea rather than fully engaging with the text and details. As a consequence, opportunities for strengthening and enriching mental vocabulary, learning from the text, and

engaging with the deeper cultural capital which is invested in these texts, may be lost. High expectations must be maintained with regard to comprehension, especially so with second language learners, most particularly because there are limited exposure opportunities compared to peers who are immersed in the dominant community language.

Reading is often described as an interactive process, not just a passive process of vocalising an interpretation of graphic images. Much of the difficult or 'stopper' vocabulary in Māori language texts is ably 'read' or more to the point 'said' with accuracy. The regular phonemic nature of the Māori language makes verbalizing text at the word level quite simple compared to some languages. However, it becomes evident in vocabulary checks that the meaning of these words often remains a mystery to the reader. As previously mentioned, sometimes making errors can intensify the interaction with the text and have a positive impact on understanding. In the long term, however, the aim of becoming a fluent reader is to gain momentum in comprehension. The less slowing down, correcting mistakes and stumbling over stopper words, the better. Readers expect the task of decoding to become less laborious as they progress and achieving a reasonable speed and good meaning make reading at the middle to upper end of schooling more pleasurable. Students move more competently from learning to read, to another level of reading to learn which places further demands on provision of materials with more content specific criteria. McDowell and Boyd (2005) identify this as an area that requires further attention, especially in regard to students moving from the junior school to the middle school. At this stage students are likely to be newly independent readers, or in transition from teacher-directed instructional reading to student-directed reading. Hancock (1999 in McDowell & Boyd 2005) says there is a risk of negative impact on student confidence that can occur for newly independent readers when faced with books that are too hard. Continued assistance in book selection for newly independent readers is important. Hancock (1999) provides strategies to support students'

transition to independent reading. These are beyond implementation in Māori-medium settings until the reading material has the necessary criteria assigned to texts, and there are more carefully organised transitional material available. May et al. (2004) assert that academic language proficiency never occurs automatically. It needs to be specifically taught. Higher level texts play a significant part in achieving this.

Another interesting feature noted was that students who had a high miscue and self-correction rate, often had good comprehension of the text. In fact, their comprehension results were often better than those of students who gave a flawless rendition, but whose comprehension results showed only minimal literal or surface understanding. The readers with high self-correction rates also used reading forward and backward strategies to sense grammatical constraints to solve a word. This presented again in their ability to strategise syntactically in this way during the vocabulary check. A student that had to stop and crack the code of a text as they went, seemed to engage more with meaning as a tool to get through the text. As a result, they employed deeper thinking in the reading process. (May et al., 2004) affirms for us that, while accuracy results for decoding texts, are useful, they are much less important than comprehension and vocabulary knowledge results. This supports the idea that measuring the lexical burden is important for educators to gauge the level of interaction and cognitive processing required in any given text. Figures 3.6, 3.7, and 3.8 compare the level of accuracy, comprehension and vocabulary knowledge attained by readers for each text in this study. The results for each group of students are shown separately. This information confirms what previous studies nationally have shown for some readers of Māori: that accuracy quickly outstrips comprehension (Berryman et al., 2001; Bishop et al., 2001; Glynn et al., 1996). Accuracy measures can create an impression of successful reading, but measures of comprehension quickly show otherwise, especially at the fluency level. This aspect of second language literacy acquisition has also been reported

internationally by Garcia (2003). Disparities between accuracy and comprehension can be compounded in second language learners of Māori because of the phonemic regularity which makes the ‘saying’ of words relatively easy. This is not the case for all students, but is significant enough to be of concern.

Figure 3.6: The Percentage scores on Accuracy, Comprehension and Vocabulary Knowledge for each Kete Pīngao i Student

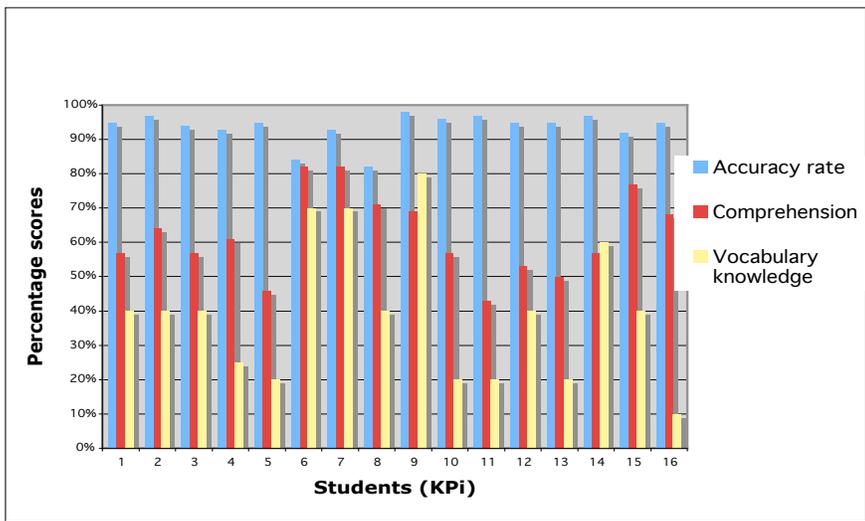


Figure 3.7: The Percentage scores on Accuracy, Comprehension and Vocabulary Knowledge for each Kete Pīngao o Student

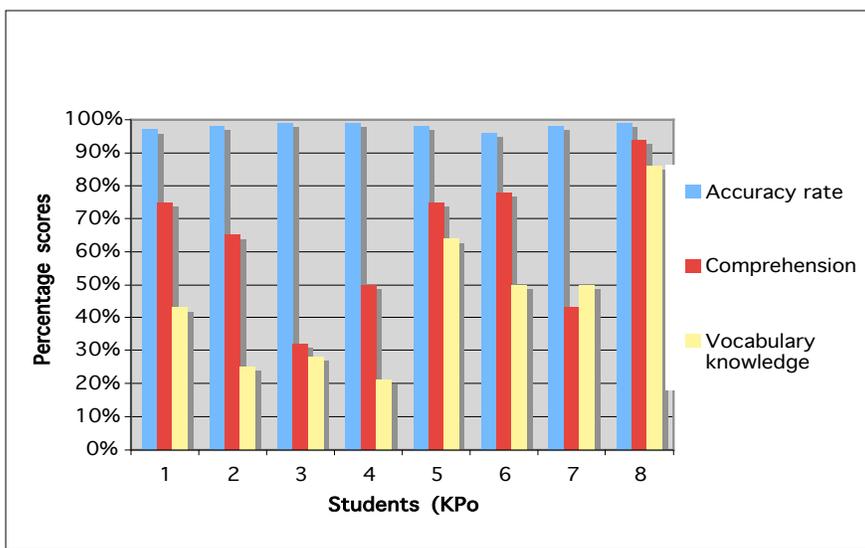
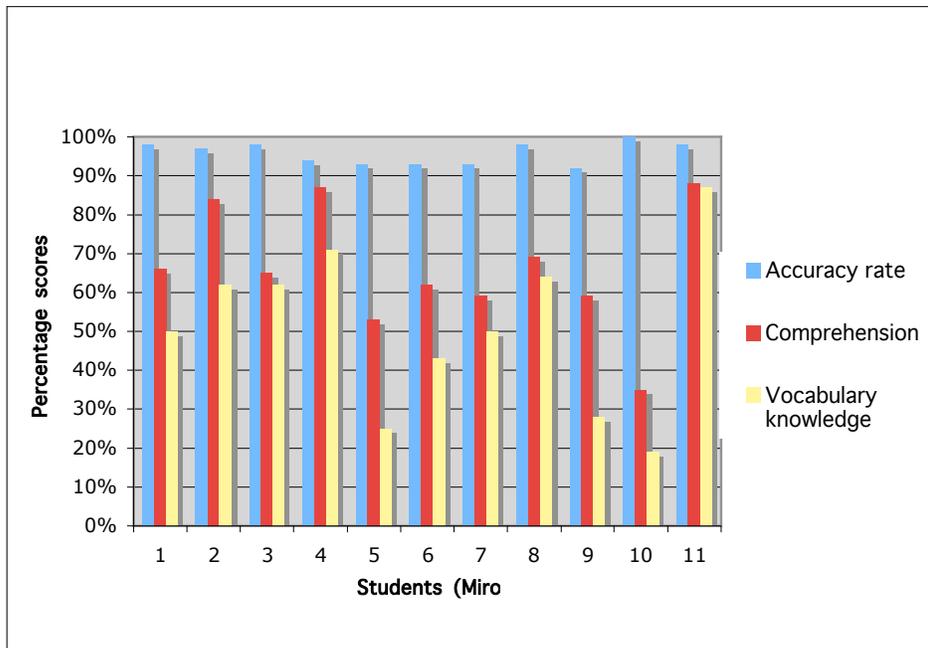


Figure 3.8: Relationship Between Accuracy, Comprehension and Vocabulary Knowledge for each Kete *Miro* Student



While all of these groups are showing high accuracy in decoding, very few students have a rate of comprehension that is keeping pace with accuracy. The general trend shown here is that lowered comprehension rates correspond with low vocabulary knowledge. There appears to be a general drag down effect connecting the two. Stahl and Fairbanks (1986) reported that instruction aimed at increasing children’s vocabulary resulted in significantly higher levels of reading comprehension. The pool of available vocabulary is an aspect of proficiency in spoken language which directly affects the level of literacy that children will achieve. This gives clear direction to educators in bilingual programmes, about the attention to vocabulary growth needed to continually support levels of comprehension.

3.5 Other findings which emerged from the study

As is usually the case, there are related aspects that appear in a study like this which are worthy of being noted and perhaps worthy of future consideration by researchers. Factors such as sentence length, illustrations, text density and syntactic complexity, reader interest and prior knowledge can also impact on readability.

3.5.1 Sentence length

Some readability formulae use sentence length as a predictor of difficulty (cf. Flesch Kincaid (1948), Fry Readability (2002), The Lexile Framework (1995)). Sentence length in these methods is used as a proxy for syntactic complexity. While not intended to be a part of this study, it was observed that some students found themselves in what Clay (1966) describes as a ‘meaning maze’ when decoding particularly long sentences. Symptoms of this occurring appeared when intonation was lost and excessive re-runs occurred. The encountering of multi-clausal sentences seemed to be the main difficulty.

An example is this sentence from the text *Kaitiaki Kaimoana*: (Waitoa, 1990) which contains 46 words. It has several related ideas about the taking of seafood. The English translation, provided to give an indication of the sentence complexity, is mine.

*Engari ia, mēnā ka iti nei te kawē i ngā kai pēnei i ngā paua māngaro,
ngā koura tika, ngā pipi, tae atu hoki ki ngā kina - kia kai tōtikatia e te
whānau – ka toe roa rā ngā kai o te pātaka o Tangaroa.*

Aside from that, if only small amounts of food such as mature pāua, legally sized crayfish, and cockles right through to sea eggs, are

harvested and consumed correctly and responsibly by the family, then the food will remain in the storehouse of Tangaroa, for a long time.

For reading material at the level of *Miro*, the increased use of multiple clauses is to be expected and is a common feature of the highly descriptive nature of modern Māori language. However, the students in this study were not very familiar with this style of writing. They had difficulty ‘suspending’ the ideas in memory until reaching the end of the sentence, at which point they knit together as a related unit. This may explain the low self-correction rate for function words in this group (see Figure 3.3). The use of multi-clausal sentences would be a worthwhile feature for those levelling texts to consider when assessing overall difficulty. Fuller research into this aspect of writing in Māori would produce better evidence as to how much the use of multiple clauses and sentence length affects understanding. Furthermore, it would be useful from a teaching perspective to have texts produced which deliberately present this feature to readers at the fluency stages. This would in turn assist in developing this skill for students to embellish their own writing and speaking. One of the key limiting factors in students’ written productive language is the capacity for sentence combining. More complex language tends to make use of the combination of several kernel sentences, where a student is required to retain longer discourse in their memory as they read. McCarthy (1954 in Barham 1965) concluded that mean sentence length in productive language was the most conclusive measure of linguistic maturity for Māori children learning English as a second language. Froese and Kurushima (1979) found that students of English understand passages written at their productive language level but have difficulty comprehending above that level. This is likely to be true for second language learners reading in Māori.

3.5.2 Illustrations

Illustrations have long been recognised as important visual clues for gaining meaning from texts. While the students in this study made extensive use of visual cues to gain meaning, this strategy, however, was not always employed successfully. In the story *Waimarie he moemoeā noaiho* (Only a dream) Nicholson (1994), one of the key characters is a 'koroheke' or old man. During the vocabulary check some students could not explain the word 'koroheke' even though many visual clues were provided about the characters. They understood there was an old man in the story because of the illustration, yet they did not connect the dialogue or the noun to the character. This shows some disconnection between vocabulary and visual clues in making good sense of the story. In another instance during the comprehension probe for the same text, a shark appearing in a 'thinking' or 'dreaming bubble' was of huge support to the readers when asked if the shark really was at the scene because the word *moemoeā* (dream) was a low frequency 'stopper' word for many. The presence of illustrative support for comprehension of the text may have interfered with the results gained for this text. For the purpose of strengthening the focus on the processing load of the text, further texts used in the research had no illustrative support.

3.5.3 Density of text and syntactic complexity

The density of the text layout in *Kaitiaki Kaimoana*, seemed to generate a high rate of line skipping. This could simply mean that this group of students were not used to such density and would overcome this after more exposure to texts like this. An interesting feature of line skipping is whether or not the reader can 'marry up' meaning from one line to the line they have skipped to. In the case of these students, reading for meaning did not secure the recognition of the error, as shown in the example below;

The skip is bracketed and we can see it has occurred at word 4 in the sentence (mai) and is picked up again at word 18 (mai).

Ka whānaunau tonu **mai** (ngā uri a te kāhui ika hei whakakī i nga whāruarua kua mahue **mai**) i ērā kua riro i te tāngata mā, hei whakakī kōpū. (p. 27).

In translation the sentence in full reads:

The young of the fish species continue to be spawned to replenish the place left in the stocks which have been taken by the people to fill their bellies.

The skipped version was read:

Ka whānaunau tonu mai i ērā kua riro i te tāngata mā, hei whakakī kōpū.

This translates as:

Spawning continues from those which have (already) been taken by the people to fill their bellies.

The skipped version reads as though the fish which have been caught and eaten are still spawning, which does not make sense. The students had lost the thread of the sentence but did not attempt to repair it. This shows that meaning at the sentence level was not being attended to. Instead, gaining a global gist of the text seemed to be their aim.

3.5.4 Reader interest and prior knowledge

The personal resources a reader has, such as, word attack skills, interest in the topic and prior knowledge were also noted. The students who showed good results in the vocabulary check and correspondingly good comprehension results were those who had prior experience and high interest in the topic and activities

in the text. This is an indisputable factor in supporting successful engagement with any text, and will always be a key element in text selection for students.

3.6 Summary

This small duplication of Marie Clay's (1966) study yielded several useful pieces of information which shaped the direction the analysis would take. This phase also assisted in taking a critical approach to transferring the Elley method onto the Māori language. In this phase of the study, the isolation of nouns as the most discriminating word class for difficulty, was not proven conclusively to be the case for the Māori language. As a result, it was decided not to proceed with a focus exclusively on nouns, but to include the wider word class of content words in the construction of the graded word lists. This phase also confirmed that other aspects need to be taken into consideration for overall readability, such as sentence length, text density, illustrative support and prior knowledge. The strong relationship between vocabulary knowledge and comprehension scores also showed that strong familiarity with content words is important for good comprehension.

Chapter Four: Phase Two Construction of the word lists

4.1 Introduction

This chapter contains a description of key approaches to the selection of words in the compilation of vocabulary lists, and details the approaches applied in this phase of the study. Elley's original noun count method (Elley 1969) was based upon a platform of word lists which were constructed from a wide range of corpus material in English. The lists were further revised using Croft's (1983c) frequency counts from a national survey of primary school children's writing. The background detailing the construction of those lists is in Elley & Croft (1989). Elley's method is one of assigning nouns a value and then generating a total vocabulary load score over a given text. The word lists used for this current study were constructed by a process of comparing and amalgamating data from the *Māori Broadcast Corpus* (Boyce, 2006), the *(Huia) Corpus of Māori Texts for Children* (Huia, in progress), *He Kupu Pū Noa* (Benton, 1982), and *Te Kura ki Uta* (Maxwell & Benton 1995). The difficulties that arose while compiling the word lists and how these have been accommodated is also explained. The word lists and their data are in Appendix 2 along with the data generated for this analysis.

4.2 Methods of compiling word lists

Most teachers of language and most providers of reading material in that language have ideas about the basic rules and vocabulary needed to establish a level of competence to operate successfully in the language. This is manifested in a wide variety of approaches to the teaching of languages and a variety of corresponding material used in the process.

Boyce, a long time teacher of the Māori language, notes that for adult learners:

One of the key factors inhibiting comprehension of the material is the proportion of vocabulary items that the learners do not know. Identifying the lexicon, and in particular the high frequency items, is therefore a positive step towards establishing priorities for learning. (2006: 13).

Vocabulary selection is a key feature of formalised language teaching, and word lists have become an important tool. The words contained in word lists are selected using subjective and objective methods or a combination of both.

Richards (1974) provides a good background on historical methods of word list construction and describes subjective and objective approaches. His conclusions support the belief that a great deal of information is contained in the low frequency words that are unfamiliar in texts when he states that:

We can recognize 80% of words in a text and yet totally fail to understand it, since the crucial information may be contained in the 'outsiders' (p72).

He describes the inclusion power of high frequency words as being important for comprehension.

Subjective methods of compiling word lists rely heavily on the intuition of very experienced teachers of the language to decide on the appropriateness and value of specific words to the learner. This is the simplest approach to word list design but it relies solely on the personal judgement of experts in the target language. Unfortunately, this type of approach often creates dissention and impressions of unreliability, as the opinions of such experts vary greatly.

In contrast, objective methods of compiling word lists propose a scientific approach to vocabulary selection. Richards (1974) notes that there are criticisms about this purely objective process, with word lists being at variance with teacher

intuition, and also because no two word lists are in substantial agreement due to the various methods employed in their production. Word frequency lists are lists of words organised by the frequency with which the words are present in large collections of written or spoken language. These collections are known as corpora, or, singularly, as a corpus.

Using computer programmes it is relatively easy to calculate both the overall frequency of a word and also how well the word is spread throughout all of the component texts in a corpus (the word's range). These two components, frequency and range, together form a picture of how often a word appears, and in how many situations it arises, indicating its likely familiarity to the reader. Fortunately, there is a good field of software available to researchers for analysing frequency and range data in large collections of text. Corpus analysis software, *Oxford WordSmithTools version 4* (Scott 2004) hereafter referred to as *WordSmith* was the primary tool used to generate frequency and range data. *Microsoft Excel* was used for storage, management, and comparison of data during the construction of the word lists. Finally, on completion of the word lists, *RANGE*, Heatley's et al. (2002) text analysis software, was used to analyse selected texts using the word lists.

4.3 Corpus materials

It was decided that the most robust approach for this analysis would be to combine both subjective and objective methods of constructing word lists. To achieve this, four corpora of the Māori language were compared, and any words showing significant discrepancies in their data were brought forward for closer analysis and teacher opinion.

4.3.1 *A Corpus of Modern Spoken Māori*, Boyce (2006)

This current study has only recently become possible since the publication of *A Corpus of Modern Spoken Māori* (Boyce, 2006). Mary Boyce has supported the pursuit of this project by making available an electronic copy of her corpus and providing assistance, including advice on computer analysis tools. *A Corpus of Modern Spoken Māori*, which contains the Māori Broadcast Corpus is a significant corpus of 273 spoken texts which contain in excess of 1,000,000 words of te reo Māori. The spoken texts were compiled from radio and television broadcasts in the Māori language in 1995. Thus the *MBC* is a representative corpus of contemporary spoken Māori. Boyce (2006) provides a detailed description of the representativeness, size, balance and quantity considered in the design of the *MBC*. The corpus is a static corpus which was collected within a finite framework and once finalised it does not change. Boyce has managed her work into database formats using *WordSmith* and the *MBC* has been published with word frequency and range data already generated. The *MBC* has not been divided into equal parts to produce range data. It has been measured using the distribution of words across the 273 texts which make up the corpus. The limitations of using the *MBC* for this phase of the study were that it does not contain the language of everyday informal conversation, nor does it contain the language of written Māori, specifically that written for children.

4.3.2 (Huia) *Corpus of Māori Texts for Children*, (Huia Publishers, ongoing)

With the permission of the Ministry of Education, further corpus material has been supplied by Huia Publishers for this project in the form of an unpublished and developing corpus of texts written for children. This is known as the *(Huia) Corpus of Māori Texts for Children*, hereafter referred to as the *CMTC*. At the time it was used in this study (2007), the *CMTC* contained 772,000 words from

1,175 texts written for children. In contrast to the static *MBC*, The *CMTC* is a dynamic corpus which is constantly changing and growing as more texts are added to it. At the time it was used for this project, it had not yet been through a 'culling' process according to a predetermined design. This means that it has some limitations because it has not yet been 'balanced' by the application of certain criteria, which is an important part of the final design of a corpus.

4.3.3 *Ko Ngā Kupu Pū Noa o Te Reo Māori, The First Basic Māori Word List. Benton, Tumoana and Robb (1982), and Basic Māori word list: levels six to ten. Benton (1983)*

These companion publications were drawn from the earliest collections of corpus material in the field of Māori education. They contain 11 graded word lists and are the only attempt to date at generating word lists using frequency data. The underlying aim of Benton's et al. (1982) research was to assist teachers to incorporate into the early stages of teaching, the most useful general purpose Māori words which are high frequency words encountered in most situations. Their study involved analysing a collection of speech and writing which contained 106,608 words of text, and assessing the value of words by looking at their frequency and representation across a range of contexts. The source material for Benton's et al. work was categorised into six domains; secondary school textbooks, wider material suitable for secondary school aged children, primary school texts, spoken recordings of native speakers of primary school age, and news and radio broadcasts intended for adults. This collection covers a good range of sources, and the 500 most valuable content words were grouped into five levels which also had 150 function words added. A further six lists of words were published the following year. The lists ascend in difficulty from levels one to ten and an additional list named *toenga*.

This current study shares the approach of identifying high and low frequency content words and grouping them into graded lists. The differences between the lists are that for this study, the function words are kept on a separate list, and the source corpora used are significantly larger.

4.3.4 *Te Kura ki Uta, Opotiki College Bilingual Education Programme Assessment and Evaluation Project, Benton & Maxwell (1995)*

Further to this early work, Maxwell and Benton in *Te Kura ki Uta*, (1995) have provided extensive information on word frequencies using a corpus of just over 160,000 running words from *Te Wharekura* and other similar classroom reading materials. This publication outlines extensive testing and also comments on vocabulary acquisition and development confirming the importance of having reading material graded. Benton says:

If learners are reading (and listening to) material which is inherently interesting and challenging but not overwhelming, they are more likely to make use of such resources and to retain new vocabulary. (1995: 5)

Benton (1995) further quotes Saville-Troike (1984: 219):

...vocabulary knowledge is the single most important area of second language (L2) competence when learning content through that language is the dependent variable.

For students in full immersion settings their competence in the Māori language is indeed the 'dependent variable'. It will determine the level of cognitive stimulation able to be provided by their formal education. Ingram and Elias (1974) warn that if a child is operating in a school system where his level of vocabulary development in the language of instruction is unable to carry sufficient stimulation, cognitive development will be compromised. There is a danger that children finding themselves in this situation are forced to retreat to a level of cognitive development inferior to that of their monolingual peers. This adds a

sense of urgency to the position that careful and effective vocabulary growth and exposure needs to be professionally managed and critically monitored for students in these settings. Māori immersion settings have often been born out of a desire for social justice, yet may risk finding themselves colluding in another institution perpetrating the linguistic and cognitive restriction that was prevalent in 19th century schooling in New Zealand (cf. Simon & Smith, 2001).

4.4. Method of prioritising data

The relevance of the four corpora to this project were prioritised with the *CMTC* being used as the key corpus to triangulate the others against. For this study, the *CMTC* was an especially useful indicator of how well-known a word could be expected to be to a student in a Māori-medium education setting, and in how many different texts it is likely to be encountered. This is because the material that makes up the *CMTC* corpus is from Ministry of Education publications distributed to Māori-medium classrooms. There are different ways to organise a corpus which in turn affect the data produced about a word's range, or its spread throughout the corpus. The range data gathered for this study left the *CMTC* corpus grouped in the original texts in order to provide an authentic picture of likely exposure across a variety of texts. This was also in line with the organization of range data produced by Boyce for the *MBC*. The frequency and range data of the *CMTC* corpus was generated using *WordSmith*. The data from the other three corpora played a pivotal comparative role across the data collection.

The importance of prioritising the *CMTC* data for this study soon became evident. For example, the *MBC* rated words such as *pūtea*, *rohe*, *whakahē*, *komiti*, *rūnanga*, *pāremata*, *kāwanatanga*, *pākehā*, and *poari* as high frequency, which indeed they are in adult speech. However, in contrast, the *CMTC* data indicated

that these words were low frequency, narrow range words in children’s texts. This point is illustrated in Table 4.1 which contrasts three different data sources for the word *poari*.

Table 4.1: Comparison of word frequency and range data from different corpora

Word	<i>MBC</i> frequency	<i>MBC</i> RANGE	<i>CMTC</i> frequency	<i>CMTC</i> RANGE	<i>He Kupu Pū Noa</i> level
poari	550	101/273	8	6/1,175	9

Table 4.1 shows that the word *poari* was used 550 times in 101 separate situations in the *MBC* while it was only used 8 times in 6 situations in the *CMTC*. In addition, Benton rated *poari* as a difficult word, putting it in level 9 out of 11 vocabulary levels.

Conversely, there were words captured by the *CMTC* frequency data and *He Kupu Pū Noa* which would not have appeared if the data were restricted to that generated from the *MBC*. Words which abound in children’s texts, for example, *kōhanga*, *tuhia*, *whārangī*, *horoi*, *peita*, *peke*, *pōro*, *pī*, *rorohiko*, *matimati*, *menemene*, *parakuihi*, *rare*, *wharekura*, and *tiakarete*, all have low frequency in the *MBC* adult speech corpus. In addition, Benton’s et al. (1982) lists did not contain words like *kōhanga*, whose usage had not become widespread at the time the list was compiled. These kinds of observations confirmed that for this analysis the *CMTC* would be the primary data source and the use of the other corpus data in a monitoring role demonstrates the rigour gained by comparing corpora. Together, the *CMTC* and Benton’s et al. data brought a balance to the *MBC* data, and provided a more specific picture of the nature of words to be found in the world of children’s texts, and their frequencies.

4.4.2 Constructing the word lists

Elley and Croft included 2050 words in their noun frequency lists for English, which they grouped into 8 score sets. Most word lists for English which are based on frequency data aim to contain approximately 2000 items of the highest frequency in the language (cf. Nation, 2001a; Elley & Croft, 1989). An analysis of the *CMTC* corpus for words which occurred 12 or more times in a minimum of 12 different texts captured approximately 2072 words, and these words comprised the base selection for the lists. In addition, the 200 most frequent content words in the *CMTC*, *MBC* and any additional words from *Te Kete Kupu* (Huia 2005) were also included in the starter list. Most of these words managed to hold their place on the lists once the cut off criteria of frequency 12 and range 12 was applied to the *CMTC*. Benton's et al. (1982/1983) collections totalling slightly more than 1,700 words were also amalgamated into the starter lists. Finally, words from the curriculum word lists (Maxwell & Benton, 1995), which were not already included in Benton's et al. lists were considered for inclusion.

There are some words for which exceptions to the criteria have been made. Words which break the parameters of frequency and range are generally due to the inclusion of Benton's et al. *He Kupu Pū Noa* (1982, 1983) collection. Benton's et al. early work brought forward some 'outsiders' for consideration, some of which were eventually included based on teacher opinion.

Function words and names were then deleted which eventually reduced the number of words on the lists to 1,827 content words.

4.4.3 Grouping the words into score sets

Using *Microsoft Excel* the data entry and grouping process was undertaken in chunks with ongoing modification as the lists took shape. The process below outlines how the starter list was progressively developed into groups.

1. Nine score sets were set up, and words were entered into a score set between 1-9. The conventions used by Elley and Benton et al. were followed, where 1 = high frequency and therefore easy, and 9 = low frequency and therefore more difficult. See Table 4.4 for the divisions of the sets based on *CMTC* frequency and range.
2. Data were entered alongside each word including fields containing:
 - *He Kupu Pū Noa* level 1-10 and toenga,
 - *CMTC* frequency,
 - *MTC* range,
 - *MBC* frequency.
 - *Te Kura Ki Uta* curriculum words not already in *He Kupu Pū Noa*
3. Exceptions were sometimes made due to the easy Benton et al. (1982) level, and teacher opinion. These words had comments added to the cells in the grouping spreadsheets for reference, stating the reason they were breaking the parameters of the score set. In total, 48 words were brought in from Benton's et al. lists based on teacher opinion and added to score sets 8 and 9. Some words were also moved because of their occurrence in some of the emergent reading texts which had not been included in the *CMTC* corpus (eg. *kurī*, *whakarongo*, and *taniwha*).

4. Words suspected of also being names were checked against concordance data from the *CMTC* corpus using *WordSmith* to ascertain how much the proper noun form had inflated the frequency. In cases where name inflation was evident, the name occurrence was subtracted from the frequency data and a comment inserted in the cell for reference. See 5.3.2 for a discussion on the inclusion of proper names.
5. Frequencies and range in the *CMTC* were added together for words which had dual entries in the corpus. These words were usually illustrative of the time where macronisation was still developing as an orthological convention; for example, *korero* and *kōrero*. Usually the macronised version is now the accepted one, although some word types may still retain non-macronisation preferences for some authors. The word is listed under the form which had the higher of the two frequencies, which in most cases is also the accepted spelling. Williams (1992) was the reference text used for standardising macronisation. Some words have two forms with a difference of only one vowel and the data for these forms were also combined, for example *taiaapa* and *taiepa*.

4.4.4 Changes in usage

Some of the words on the lists reflect a rise in usage of words driven by curriculum development. These are words which would not have made it onto the lists if only Benton's et al. lists and the *MBC* were used. For example, the word *taumata* (used in Māori curricula for 'level') along with other curriculum focussed words like *pūtaiao*, *hauora*, *hangarau*, *toi*, *pāngarau*, *hākinakina*, *rorohiko*, *kauwhata*, *kiromita*, *tapawhā* and *ine* which do not appear in Benton's et al. lists or are very low frequency in the *MBC*. In contrast the *CMTC* data shows common usage of these words especially in texts related to mathematics.

Conversely, some words have become dated and have decreased in usage. For example *māhita* (for teacher, derived from ‘master’) has been replaced with other words like *kaiako*, and *pouako*. Other borrowings from English in Benton’s lists have also become less preferred like *tiaka* (jug), *eroperina* (airplane), *kiki* (kick), and *tīma* (boat). However, older speakers still regularly use some of these earlier borrowed words as evidenced by their frequencies in the adult spoken corpus of the *MBC*.

Changes like this over time illustrate the natural and evolutionary process of language change. It reinforces the need for vocabulary lists and text levelling processes to be continually monitored so these changes can be incorporated. Elley and Croft revised their 1975 lists in 1989, removing dated words, adding new words, and changing placements to different levels.

4.4.5 Months of the year, days of the week and numbers 1-10

The data from the corpora placed words for months, days and numbers in a variety of score sets according to frequency and range. Like Benton and Elley, these words were treated as a separate group because these words get regular exposure in classroom settings although they do not often find their way into written texts. Teacher opinion assisted with the placement of these words.

The words for numbers one to ten were placed into score set 1 because they are regularly encountered in written form by children right from their entry to school. The words for numbers one, two, three and ten (*tahi*, *rua*, *toru*, and *tekau*), landed in set 1 by their frequency alone, the others were added.

The borrowed words for days of the week and their more recently coined counterparts were all placed into score set 2 on the basis that these are words

which are used in classrooms on a daily basis even though they may not occur frequently in texts written for children.

Some words for months of the year did not make it on to the lists at all, and these words are less frequent in daily writing than the days of the week. The placement of these words was decided by looking at where the majority of them fell according to their own frequency data. As a result of this, the borrowed words for months of the year are all placed in score set 8, while the more recently revived months were placed in score set 7. This latter set of words for the months has risen in prevalence in texts and spoken usage in recent times. This allowance was made because it was considered that most children experience these words every day for a month of the year in written form as a regular part of classroom routines.

4.4.6 Inclusion of word lists from *Te Kura Ki Uta*

In 1995 Ian Maxwell and Richard Benton compiled further word lists as a part of *Te Kura ki Uta, the Opotiki Bilingual Education Evaluation Project* (Maxwell & Benton, 1995). These lists were compiled from Benton's earlier lists in *He Kupu Pū Noa* and also included the word lists in the appendices of the *Te Reo Māori Marautanga* document (Ministry of Education, 1996). In order to make the construction of the word lists for this study as thorough as possible for the educational context, it was decided to incorporate the curriculum lists into the data. This was done primarily to provide extra screening to the data in the *CMTC* but it also revealed how much of the recommended curriculum vocabulary actually appeared in Ministry publications. The analysis revealed that 357 words out of 2,680 which are listed in the appendices of the curriculum document *Te Reo Māori Marautanga* (Ministry of Education, 1996) as Māori vocabulary to be experienced by children across the curriculum, never appeared once in any of the publications incorporated in the *CMTC* corpus, and very rarely in the *MBC*. A

further 205 words had fewer than six occurrences. This showed that much of the vocabulary incorporated in the curriculum appendices at the time were not later supported into curriculum print material. Even the lists at koeke 1-3 (levels 1-3) had 228 words out of 1,612 that appeared fewer than six times in published materials, and 78 of those words did not appear at all. The curriculum has now been reviewed and a new curriculum is in the consultation stages. This shows that there needs to be better follow up on making sure that if vocabulary lists for curriculum levels are going to be recommended, that material written for use by children in classrooms includes that vocabulary.

4.4.7 Assigning values according to word frequency

The task of assigning values according to word frequency was guided by the work of Nation & Worthington (1996), Elley & Croft (1989), and Benton et al. (1982). Finding a way to rate the words most likely to be unknown to the reader is believed to give an indication of the challenge a text will pose. This can be achieved by analysing word frequency counts.

Most frequency word lists work with collections of around 2000 words. Elley & Croft (1989) and Benton (1983) divided their collections into around nine or ten groups, although Maxwell and Benton (1995) divided lists into just three groups. Criteria used to define where the cut off points go, vary depending on the purpose and application of the lists. Although determining just where to make the cuts between one frequency band and the next are somewhat arbitrary, Nation and Worthington (1996: 1) say that:

Although the dividing line between high frequency wide range vocabulary and low frequency narrow range vocabulary is largely arbitrarily drawn, there are striking differences in the return to the learner for the effort in learning these two types.

The distinctions between the cut off for scoring the words in this current study were decided by naturally occurring gaps in the frequencies. These gaps kept the lists within 100 or so words of each other, with more words contained in the high scoring low frequency bands. The Elley & Croft lists show similar patterns. The three other main sources of data (*MBC*, *He Kupu Pū Noa* and *Te Kura Ki Uta*) have sat alongside the words to alert the researcher to any atypical frequencies that the *CMTC* may contain while still in its unfinished form. An example of how this system worked is given by the word '*taki*'. The *CMTC* frequency of 298 and range of 222 placed it in score 2 (easy). However Benton had it in level 10 (hard) and in the *MBC* it only had a frequency of 50. This disparity alerted the researcher to check the use of the word in the *CMTC*. It was found that '*taki*' was inflated in the corpus by use in a name, *Te Pou Taki Kōrero* (Learning Media) and once the proper noun was removed, the word was moved from the score 2 list up in difficulty to the score 7 list. This process of following up such alerts was an ongoing task during the construction of the lists.

4.4.8 Word sense

For Māori, as in other languages, there are word forms that are polysemous, or carry multiple meanings. These need careful consideration when constructing word lists as mentioned in Maxwell & Benton (1995). There are also older Māori words that exist alongside recent borrowings. An example of one such word found in the Benton et al. (1982) lists is *whiti*, to shine, which through borrowing, is listed again as a separate word form and given an English meaning of 'fit'. Another example is *tari* used as a noun in a modern sense for office, borrowed because it sounds like *study*, but as an older word, has the meaning of the verb *carry*. Therefore, word sense can interfere with the general application of frequency data. Some corpus software has been developed to grammatically tag and parse words into their grammatical class. Unfortunately this is not yet

available for Māori language. This means that the only way to overcome issues relating to word sense is to visit each entry manually and by using context and collocates, decide on the meaning and separate out the frequencies. While a small amount of this was undertaken especially with regard to names, it was not possible within the scope of this analysis to do this for all words that presented this difficulty. Benton (1982a) explains how this complication was managed in the construction of his early word lists for Māori. Benton's et al. entries of different word forms with the same meanings were incorporated into this analysis by using only the true occurrences of words.

4.4.9 Word families

Another issue that presents itself in the construction of word lists is that of word families. A word family consists of a base word and all of its derived and inflected forms. This includes suffixes and prefixes. Worthington and Nation (1996: 1) describe a word family as “a base word, for example, dig, and closely related inflected forms and derived forms; digs, dug, digging and digger”. A familiar example in the Māori language is a base word like *haere*. It can have many members in the family, such as, *haerenga*, *haeretia*, *haeretanga*, *haerengia*, *hāereere*, *whakahaere*, and *kaiwhakahaere*. Without giving a detailed glossary of these words, one can identify the base word *haere* occurring throughout this family. While this aspect does increase the base word's value, the difficulty that the prefixed, suffixed, or derived form will present to a reader unfamiliar with its written form, remains unknown. Bauer and Nation (1993) discuss methods of dealing with inflected and affixed forms of words and present a practical system for levelling affixed forms in English. They produced a graded set of seven levels of difficulty to be considered when using these forms in levelled texts. Word frequency remained their primary criterion. Bauer and Nation describe the complexity of relational word knowledge required of native speakers to cope with

a range of word families in text. This is likely to be far less developed for a learner of a second language. To date there have been no studies undertaken for children reading Māori, that determine whether or not there is added comprehension burden by the appearance of inflected and derived forms in reading texts. However, it may be reasonable to assume that the complexities that these have been found to present for other languages may also be true for Māori. Affixed words in this project were sorted into word lists by their frequency of occurrence and not by family. In other words, they were treated as separate content words.

Some of the difficulties presented by word families for this study can be observed in the differences in the frequency and range between the base word and the inflected forms. Some examples of this are presented in Table 4.2 which shows some examples of discrepancies in frequency and range that word families can present. (Keep in mind the number of contributing texts in each corpus, *CMTC* = 1,175 / *MBC* = 273).

Table 4.2: Examples of difference in frequency and range between bases and their inflected forms

Word	<i>CMTC</i> frequency	<i>CMTC</i> range	<i>MBC</i> frequency	<i>MBC</i> range
āhua	1,359	419	1,537	245
āhuatanga	392	205	1,631	245
haere	4,154	676	5,556	273
haerenga	236	129	168	98
kōrero	3,416	882	6,625	273
kōrerotia	55	47	215	108

It can be seen in Table 4.2 that in written texts from the *CMTC* the root form is generally more frequent than the derived or inflected form. However, in the *MBC* for example, the derived form of āhuatanga is slightly higher than the root form

āhua and both forms have the same range of occurrence. Bauer & Nation (1993) say that criteria involving frequency, regularity, productivity and predictability need to be taken into account when rating difficulty of word family members.

Sometimes reduplicated forms of words have higher frequencies than their root form. Some examples of this are shown in Table 4.3.

Table 4.3: Examples of different frequencies of reduplications and the root word forms

Word	<i>CMTC</i> frequency	<i>CMTC</i> range	<i>MBC</i> Frequency	<i>MBC</i> range
kōpikopiko	27	15	5	6
kōpiko	7	5	3	1
mirimiri	48	33	13	6
miri	29	12	6	6
katakata	81	53	20	16
kata	259	112	42	23
purapura	17	7	17	8
pura	6	4	3	3
waewae	366	170	140	67
wae	25	10	9	4

Table 4.3 shows that a process of assigning the root word the easiest rating (ie highest frequency) is not always reliable. Conversely, the case of *kata* is one instance where the root word does have the highest frequency, even though one intuitively might not have expected this to be the case. Due to the features that the inflected and reduplicated forms present to the reader, with increased word length and syllabification, it was decided to treat these as separate items for scoring. Because there is no research into the impact word families have on reading in Māori, score ratings were made primarily according to the *CMTC*

data and not on the basis of belonging to word families. Further study into this aspect would inform the revision of the lists in this analysis.

Other word lists that have been constructed on the basis of frequency alone have found ways to distinguish between words belonging to the same word family while still grouping them together. See, for example, the NCEA word lists for English as explained by Wallace (2003) where each member in a word family was analysed for frequency and given a star rating based upon the number of occurrences in the corpus. The words belonging in the family are listed together with their individual star rating.

The star frequency rating system used in the NCEA word list is as follows:

FREQUENCY IN THE NCEA CORPUS:	
Star Rating	Occurrence of word
*****	200 or more occurrences
****	100 – 199
***	50 – 99
**	20 - 49
*	1 - 20
(no stars)	does not occur

For example, the final entry for the word family of 'interpret' appears as follows:

INTERPRET v ***
INTERPRETATION n***
INTERPRETATIVE adj. *

Although a word's range of occurrence does not appear to have been included, this method of grouping words is similar to what has been applied in this analysis except that the words are placed in different score sets rather than being star rated and in families.

4.4.10 Plural forms

This study followed the same process of dealing with plurals as Elley & Croft (1989), which is to combine the placement of the singular and plural form. While, in general, most nouns in Māori retain their form in both singular and plural contexts, there is a small collection of kinship words in which the antepenultimate vowel is reduplicated. This generates the plural which is signalled orthographically by the use of a macron. Plural forms have been placed in a score set with the singular form. In every case the singular form has the higher frequency. Examples of these word pairs are: *wahine*, *wāhine*; *tangata*, *tāngata*; *teina*, *tēina*; *taina*, *tāina*; *tipuna*, *tīpuna*; *tupuna*, *tūpuna*.

4.4.11 Word utility

Richards (1974) suggests that teacher opinion regarding the concrete nature of some words (as opposed to words for abstract ideas), could be incorporated into decisions about placement of words onto lists. Richards also discusses the value of a word, or its utility, meaning how useful the word might be in everyday life and how transferable across contexts a word might be. These aspects were briefly touched on for this analysis, with words such as *wharepaku* (toilet), where teacher opinion promoted its score based upon the word's frequency of use and high familiarity in spoken situations. Although it did not have high frequency data in the corpus material, it has high print exposure in school and public settings through signage. There are other words not on the lists such as *whakamate* (to kill) which are prevalent in the conversation of children which come directly from recounting children's television programmes. Other words which did not make it onto the word lists also require further consideration. Words such as *hanawiti* (sandwich), *kapahaka* (traditional dance), *kōwhaiwhai* (traditional art pattern), and *ipupara* (rubbish bin). These are all well-known Māori words which are

spoken in school settings, yet they are not written about often enough to gain a place on the lists. The scope of this analysis does not permit a study into how or where to place such words. Omissions of words like this demonstrates the need for a corpus of language spoken and written by children.

4.4.12 Function word lists

In addition to the content word lists, a list of function words was made and contributes an added dimension to this study. This measure was adopted following a recommendation made by Benton et al. (1995) which suggests that a simple measure of syntactical complexity can be estimated by counting the number of function words used in a text. In addition, Peter Keegan (personal communication) at the University of Auckland separately recommended some simple proxy to measure syntactic load. Another influence in generating this extra list was the software program *RANGE* (Heatley et al. 2002) which was used to analyse the texts and has a function word list in the English version. This feature separates out function words from a text which proved very useful for computing a simple count of the number of different function types appearing in a text. In order to use this facility, Benton's et al. list of function words (1995) was entered as a starter list and then gradually added to as more function types arose in a variety of texts. Variant spellings and dialectal forms were also included, for example, *engari* and *erangi*. There are a few problematic words which are used as both content and function words, for example *hoki*, *mātau*, *mei*, *wai*, *rā*. The decision about which type of list to place them on was determined by examining their highest use in the *CMTC* corpus. Once the software program *RANGE* began running over the texts even more words came up for incorporation into this list. The incorporation of this function word list into the final calculations is explained further in Chapter 5. The list of function words is in Appendix 3.

4.5 Developing the Baseword Lists for *RANGE*

Software programs for analysing vocabulary have been developed by the English Language Institute at Victoria University over several years, these have included *VORDS* and *FVORDS* which were superceded by *VocabProfile* which has now been developed further. *RANGE* is the latest version of this software and was designed at the Victoria University of Wellington by Averil Coxhead and Paul Nation, a leading vocabulary researcher, and was programmed by Alex Heatley. This program was developed to assist vocabulary analysis of texts in English. It requires a set of lists known as ‘baseword lists’ to analyse texts against. The program analyses a text or group of texts and identifies which of the baseword lists the words in the texts come from. In addition, it produces a list of words that do not appear on any of the baseword lists. It was possible to customise Māori language baseword lists for *RANGE* to use with a few extra modifications for macrons and phonemic structure. This program proved to be a very reliable and effective way to analyse the selected texts.

Ten baseword lists were prepared in total: nine lists of content words based on frequency and range values and one list of function words. The data generated by the *RANGE* analysis was used to select texts to trial with teachers and students. The analysis process also contributed to the robustness of the word lists as it constantly profiled variant spellings, macronisation and typeset errors in texts. This also accounts for the differing number of words on the word lists and the Baseword Lists. Chapter 5 details how *RANGE* was used to select and rank the texts.

4.6 Word list totals

Table 4.4 shows the total number of words in each of the 10 word lists. It also shows the cut off criteria for frequency and range of occurrence. The overlap in

frequency cut off between lists occurs because of the range data being applied. For example the frequency band for words on lists 4 and 5 overlap. However, a word occurring 50 times in 50 different texts must be considered more likely to be met by readers than a word that occurs 50 times but only in 20 texts. This is because some words are used a lot only in a specific context or by a particular author. The frequency data from the corpora just happened to present a large group of words with frequencies between 220 and 50 but which had a wide range of occurrence across texts. This group of words were therefore more accurately delineated by looking at their range of occurrence. This shows how the interface of frequency and range data was applied together when drawing up the lists and deciding on cut off criteria. As previously mentioned, by incorporating teacher opinion, some exceptions to the criteria were allowed. Some words are included in lists even though they did not fall within the exact boundaries of that list. The reasons for this are explained below.

Table 4.4: Word list totals, frequency and range criteria, and number of exceptions

Word list (also the score set)	Number of words in score set	Frequency across 775,000 running words	Range across 1175 texts	Number of exceptions added to the list
1	134	400 +	100+	3
2	147	400 - 200	60 +	27
3	168	199 - 100	60+	9
4	231	220 - 50	41- 59	6
5	216	170 -50	20 - 40	5
6	178	40 - 120	12 +	0
7	235	30 - 39	12 +	9
8	275	20 - 29	11+	12
9	236	12 - 19	11 +	39
Total content words	1820			
Function word list (Frequency and range were not applied to this list)	157			
Total words	1977			

The exceptions to the cut off criteria were predominantly due to the incorporation of teacher opinion which was prompted by corpora data that did not concur. The high number of exceptions in list 2 is due to the inclusion of 14 names for days of the week and other words that were rated easy by Benton et al. (1982). Most of these were already on lists but were moved to easier score sets. Exceptions in lists seven and eight comprise mainly months of the year. The high number of exceptions for list nine is due to the inclusion of a number of words which are known to have high use in school settings, or were on Benton's lists but did not make the cut off mark using the *CMTC*. For example, the word *māhita* (teacher, master) is a word with high frequency in the *MBC*, was rated 2 (easy) by Benton et al., and is often still used in school settings. Therefore, the *CMTC* data for *māhita* meant it would not make the frequency or range cut off but teacher opinion recommended its inclusion on the lists. These exceptions which have been added to the lists appear in bold type in the word lists in Appendix 2.

4.6.1 Dialectal and spelling variance

As texts were processed through *RANGE*, some of the words appearing as 'not on the lists' were actually types that were on the lists somewhere, but with a minor dialectal or spelling variation. Because the texts being processed have been published over a wide period of time, there are differences in some of the accepted orthography of the time. This spelling variance is what necessitated the addition of some words into the Baseword Lists, but more importantly, this process accommodated dialectal difference into the Baseword Lists. Table 4.5 shows the total number of words on each word list, and the number of differences added, which became the Baseword Lists for *RANGE*.

Table 4.5: Dialectal and spelling variance between score set totals and Baseword Lists

Word list (also score set)	Number of words in score set	Dialect / Spelling variances in <i>RANGE</i> Baseword List	Number of words in <i>RANGE</i> Baseword List
1	134	11	145
2	147	12	159
3	168	11	179
4	231	15	246
5	216	8	224
6	178	14	192
7	235	5	240
8	275	13	288
9	236	13	249
Total no. of content words	1820	101	1922
Function words (list 10)	157	100	257
Total	1977	201	2,179

The actual dialectal and spelling variances accommodated in each list are in Appendix 4. They have been included because of differences as they have appeared in texts. They are by no means an exhaustive collection and will be added to as more texts are processed through *RANGE*, alternatively the texts would need to have their spelling standardised. The *RANGE* programme also has a function that allows words to be grouped into word families. Further research in the area of word families could be accommodated using this function.

Summary

The process of amalgamating large sources of data to create the word lists and assign score sets was a very time consuming task to undertake even with the help of computer programmes. The task was further enlarged by the triangulation process of collecting and incorporating teacher opinion. It has, however, been a

worthwhile process to ensure that the construction of the lists consulted all major sources of information relating to frequency, range and also anecdotal information about words frequently used by children. The spreading of the words into word lists, then into score sets and finally into baseword lists for the computer program to use, was a process which necessitated ongoing monitoring of word placement. The final result has seen a thorough application of the combined objective and subjective approach to constructing word lists. The word lists are now developed to the stage that they can be applied to measure the semantic load and syntactic complexity of texts. This is the next phase which is presented in the following chapter.

Chapter Five: Phase Three

Linguistic analysis, ranking and selection of text series using word frequency

5.1 Introduction

In this chapter the word lists generated for this phase of the study were applied to rank and select two series of texts. The series used for this study were selected using modifications of Elley's word frequency count method for establishing vocabulary burden. These modifications, using the software *RANGE* and the calculations used in the process of selecting texts for the series are explained. A brief comparison of Benton's et al. work on levelling material is also presented.

5.2 Using the software program *RANGE*

The program *RANGE* (Heatley et al. 2002) uses up to 10 baseword lists to group words in a text or collection of texts by identifying which of the baseword lists the word is in. Baseword lists can be compiled using single words or words arranged in word families. This study used single words. The 10 word lists that were developed in Phase Two were formatted into baseword lists for use with *RANGE*. The Baseword Lists also had added to them any variations in spelling that arose during analyses of texts. These variations in spelling account for the differences in the number of words on the word frequency lists and those on the baseword lists made for *RANGE* to use. This accommodation needed to be made because unlike the *MBC*, the *CMTC* corpus has not had its spelling standardised. For this study *RANGE* was using 9 baseword lists of content words and 1 baseword list of function words.

The Elley method uses passages from a text that need to be long enough to contain at least 25 different nouns. Most of the texts used in Elley's trials

contained 150-200 words, usually segments of text. When trialling *RANGE* initially for this phase, segments of texts each containing 200 words were analysed. However, after comparing the results from using segments against results from using the whole text, it was decided that the whole text would be processed. In the 1960s, texts had to be analysed manually which was made manageable by using segments of texts. Nowadays, with computer software, it is easier and more thorough to analyse the whole text. This reveals the total measure of vocabulary load, and also eliminates sampling error.

5.2.1 *RANGE* output data

RANGE produces data for a text which shows totals for words which appear in the text from each of the Baseword Lists. It also lists any words which were used in the text but are not on any of the Baseword Lists. Tables 5.1a & 5.1b shows a *RANGE* analysis of a short text *Hōhepa te Pūru* (Yates 1986), and how the words become grouped into lists.

Table 5.1a: *RANGE* output data showing actual words ('types') occurring for the text *Hōhepa te Pūru* grouped into the Baseword Lists

<u>BASE ONE</u> hoa mea pātai titiro haere oma hoki kaha kite noho whai	<u>BASE FOUR</u> mokemoke <u>BASE FIVE</u> hipi <u>BASE SIX</u> <u>BASE SEVEN</u> pūru mangu	<u>BASE TEN</u> ka a he ko hei ia koe mai mōku aueee i te e mōu ahau arā au	engari kāhore kāo kātahi me o tēnei āe ōna <u>Types Not Found In Any List</u> Hōhepa Merenia Nanenane eeee nane ūmere
<u>BASE TWO</u> kau ātaahua	<u>BASE EIGHT</u> rakiraki		
<u>BASE THREE</u> poaka	<u>BASE NINE</u> tameheihei		

Table 5.1a gives a quick overview for teachers of the vocabulary items that a reader will be faced with in this text, and provides data which are easily transferred into the vocabulary load calculations for the methods used in this study.

Table 5.1b: *RANGE* output data for occurrences in word lists for *Hōhepa te Pūru*

Baseword List	Tokens (number of words)	% of text covered by tokens	Types (number of different words)	% of text covered by types
one	41	19.81	11	21.15
two	2	0.97	2	3.85
three	3	1.45	1	1.92
four	2	0.97	1	1.92
five	3	1.45	1	1.92
six	0	0.00	0	0.00
seven	8	3.86	2	3.85
eight	3	1.45	1	1.92
nine	3	1.45	1	1.92
ten	124	59.90	26	50.00
not in the lists	18	8.70	6	11.54
Total	207	100	52	100

Table 5.1b shows us that this text contained a total of 207 running words or ‘tokens’. Of those 207 words, 52 ‘types’ or different words occurred. According to the Baseword Lists used for this study, 41 words were from Baseword List 1 using 11 types. These 11 word types used 41 times make up 21.15% of this text. As expected, the majority of words came from Baseword list 10; 26 word types within 124 tokens, which covers 59.90% of tokens and 50% of the types. The pattern of high frequency words from Baseword Lists 1 and 10 providing the highest coverage was common to most texts.

5.2.2 Analysing text coverage using *RANGE*

Text coverage is described by Laufer (1997) as the actual percentage of words in a text which will be understood by a reader. Worthington and Nation (1996) say

that for most written texts in English, the 2,000 word high frequency vocabulary on the West (1953) list accounts for around 80% of the total running words (tokens in the text), and even higher rates of 90% for texts written for young native speakers of English. Research has not yet been undertaken to find the coverage provided by the first 1000 or 2000 words found in most written texts in Māori.

As seen in Table 5.2 *RANGE* can show how much lexical coverage from a text that each of the ten Baseword Lists provides. The example below is from *Kurī Heahea* (Gillet 1996).

Table 5.2: *RANGE* output data of occurrences in Baseword Lists for *Kurī Heahea*

Baseword List	Tokens (number of words)	% of text covered by tokens	Types (number of different words)	% of text covered by types
one	245	17.99	64	22.70
two	53	3.89	30	10.64
three	54	3.96	23	8.16
four	27	1.98	22	7.80
five	12	0.88	8	2.84
six	7	0.51	6	2.13
seven	25	1.84	12	4.26
eight	7	0.51	3	1.06
nine	2	0.15	2	0.71
ten	863	63.36	88	31.21
not in the lists	67	4.92	24	8.51
Total	1362	100	282	100

This text had a total of 1362 running words (tokens). Table 5.2 shows that of those 1362 words, 282 different words (types) occurred. According to the Baseword Lists used for this study, 64 word types occurred from Baseword List 1. These 64 types occurred 245 times which makes up 17.99% of the total running words in the text. Therefore knowing these 64 word types will give you

17.99% coverage of this text. According to the data used in this study, Baseword Lists 1 to 5 and Baseword List 10 (the highest frequency lists) added together contain the most frequent 1053 words used in Māori. By combining the percentage data of text coverage from Baseword Lists 1, 2, 3, 4, 5 and 10, we can see the overall coverage a reader will gain from this text if they know slightly more than the most frequent 1000 words used in Māori. That word knowledge would provide a reader with 92% coverage of all of the words encountered in this text. Appendix 5 shows that for the higher level texts that were eventually analysed for this study, an average of 92% coverage is provided by the most frequent 1053 words in Māori with a range of 81% - 96%

5.3 Vocabulary Index: Method M1

Elley's noun count method for calculating a vocabulary load was modified for use in this analysis. The key difference is the use of frequency lists compiled from all open word classes as opposed to just nouns. Furthermore, it includes a score for proper names into the calculations for the vocabulary load index. Reasons for this are discussed in section 5.3.2.

The nine lists of content words created for this study have been assigned a corresponding vocabulary load score. That is, all of the words in List 1 (easy, well-known words) are given a score rating of 1. Words in List 2 have a score rating of 2, and so on, up to words in List 9 (harder, lesser known words) which score a 9. Words that are not included in any of the lists score a 10 because they did not appear in this collection of the 1820 most frequently used content words in texts written in Māori for children. The function words in List 10 only score a 1 because they are generally amongst the most frequently occurring words. List 10 as shown in Tables 5.1 and 5.2 is expected to have the highest number of tokens because these grammatical words occur very frequently in Māori texts of all types. It is also expected that the content words from Lists 1, 2 and 3 will

represent the highest number of types. These are the most frequently used content words as identified by the corpora data used to construct the lists. As texts begin to use more words from the higher scoring lists, the text will pose more challenge to the reader, and will accordingly carry a higher vocabulary load score. Texts with a high number of words not on the lists will theoretically be posing a high challenge for readers.

5.3.1 Calculating a vocabulary load index

As in the Elley method, words are only counted once (by type). The words occurring in a text, are sorted into score sets (using *RANGE and Excel*) and have their scores totalled. For example, 8 words occurring in score set 3 (8x3) will contribute 24 towards the total raw vocabulary load score. Therefore, with the text used in the previous Table 5.2 as an example, Table 5.3 shows how the raw vocabulary load score total is arrived at. Finally, to account for differing text lengths and correspondingly, differing numbers of content word types possible for scoring, the raw vocabulary score is divided by the total number of content types to find the average of the raw vocabulary score for the text. This now becomes known as the Vocabulary Index (VI). For the example in Table 5.3, the VI of 3.73 as calculated for this text, can be entered alongside the VI scores of other texts and gives an idea of vocabulary load the text carry in comparison to each other.

Table 5.3: Vocabulary index score calculations (M1) for *Kurī Heahea*

List name	List one	List two	List three	List four	List five	List six	List seven	List eight	List nine	Not on lists	Raw vocab score	Total content types
Number of word types multiplied by list score	64 x1	30 x2	23 x3	22 x4	8 x5	6 x6	12 x7	3 x8	2 x9	24 x10		194
Weighted Total	64	60	69	88	40	36	84	24	18	240	723	

Raw vocab score ÷ number of content types= VI

$$723 \div 194 = 3.73$$

A series of six texts were ranked and selected using this method (labelled M1). This became known as the ‘Blue series’ of texts see section 5.7.1.

5.3.2 Inclusion of proper names

Studies for most other languages consistently find that 95-98% of text coverage is required to achieve a reasonable level of understanding. However, in these studies proper names are counted as known items. This exclusion of proper names from calculations of difficulty assumes they will be known vocabulary items and, therefore, pose insignificant cognitive burden. In a study by Hancioglu and Eldrige (2007) using similar text analysis software to *RANGE*, proper nouns were put into the off-list category during the selection of texts for reader ranking. These names were then reclassified as high frequency only if they considered them to be well-known to their participant group. Ghadirian (2002) allows the selector of texts in the text ladder computational levelling system to choose an option to include or exclude proper nouns and proper names, reasoning that

evidence is too inconclusive about the role these play in processing burden. Kobeleva believes that not enough attention has been paid to the difficulty that proper names in a text can pose. She assigns three levels of cognitive processing that the occurrence of proper names can present:

Firstly, a learner has to establish whether the lexical item they have encountered is a proper name or an ordinary expression (recognition level). Secondly, they must work out what kind of referent this proper name refers to (categorization level). Thirdly, sometimes the text will draw on additional extralinguistic information about the characteristics of some particular referent (referent properties level). (2005: 28).

Discussions with teachers, and the feedback from questionnaires on each text (see Chapter 9), concurred that the presence of names does pose a difficulty. Teachers said that the more names in a text and their associated syntactic connections of inter-relationship, the greater the difficulty. In many cases in Māori texts, common nouns and adjectives are transformed into use as proper names in the text. This is partly due to the metaphorical style used often in the Māori language. For example, one text used *nui*, a very high frequency item of common vocabulary, as the name for the main character in the narrative. This can cause difficulty for newly independent readers who often do not see the automatic salience that capitalisation carries in signalling a proper name.

In another of the texts analysed in this study, *Te Nawe a Ngā Rākau* (Fitzgerald 1998), ten proper nouns for native trees were used as proper names for characters. Teachers recorded this in the questionnaire for this text as posing particular difficulty for comprehension. Students also commented on this. Furthermore, for this text, extralinguistic knowledge was of great benefit in gaining understanding. A reader who knew in which part of the landscape particular trees generally grew, had an advantage in understanding the likely location and movement of the trees in the story.

In the light of this uncertainty it was decided that until more is known about the impact of proper names used in texts written in Māori for children, that proper names would carry the same load as unknown items or words not on the lists; that is, a score of 10.

5.4 Vocabulary Index multiplied by function types: Method M2

This method of rating text difficulty is one that developed as the research progressed. Benton et al. (1995) suggested that a good rule of thumb to measure the likely difficulty of a text in Māori is to count the number of different function words used in the text. If a number of different grammatical constructions are more tightly packed into a text, its structure is more difficult to decode for a language learner. This represents a syntactic proxy or ‘stand in’ for showing how complex the sentence structure might be, and hence provides a structural variable. Furthermore, the higher the number of pronouns in a text, the higher the chance of referential cohesion problems for the reader. In contrast, readability research for English and second language learning contains strong debate about the role that syntax plays in comprehension. Laufer (1997) quotes Ulijn and Strother (1990: 38) saying that: “while a complete conceptual and lexical analysis may be necessary for reading comprehension, a thorough syntactic analysis is not.” Elley & Croft (1989) say that adding a rating of function words for readability analysis adds considerably to the laboriousness of the process without any noticeable increase in validity. While this may have been the case for manually applying the noun count method to a text, using *RANGE* and *Excel* together makes this combination manageable. Table 5.4 illustrates an example of how the addition of a syntactic measure gives a broader picture of likely text difficulty.

Table 5.4: Texts showing differences in vocabulary load and syntactic load

NKK level	Text	Number of content types	Raw vocab score	Vocab Index	No. of Function types	Vocab Index x function types
Harakeke e	He kurī	7	12	1.71	3	5.14
Harakeke i	Ngā manu i runga i te rākau	14	14	1.00	7	7.00

As shown in Table 5.4 the text *He kurī* (Hunia, 1984) has a Vocabulary Index of 1.71 which, when multiplied by 3 function types, gives it a combined measure of 5.14. However, while the text *Ngā manu i runga i te rākau* (Gillies, 1984) has a lower Vocabulary Index than *He kurī*, once the difference in the number of function words to be processed by the reader is incorporated, we see that *Ngā manu i runga i te rākau* is slightly more complex. Multiplying the vocabulary index by the number of function types is an attempt to transform the proxies of word frequency counts and function variety into semantic load and syntactic complexity and even out the effects these have within and between texts. This simple example shows the idea behind combining a lexical variable and a structural variable. It was decided that this syntactic proxy would be combined with the vocabulary index method to generate a second series of texts. This measure is referred to as M2. This set was selected by multiplying the Vocabulary Index by the number of function types occurring in the text and was named the ‘Orange series’ of texts (see section 5.7.2).

5.5 Ranking previously levelled texts using *RANGE*

To gain an early indication of how well these two methods agreed with the early levelling of texts, a preliminary set of texts were analysed using the *RANGE* program. This set has one text from each of the 10 levels determined by Benton et al. (1995). These levels were previously described in section 2.4. The eleventh

level known as *Whatu* had no texts assigned. The texts used in this collection are from the *Ngā Kete Kōrero* (NKK) and *He Purapura* (HP) series of readers. Table 5.5 shows the titles from the 10 levels which were used for this correlation.

Table 5.5: Texts levelled by Benton et al. (1995) using the *Ngā Kete Kōrero Framework*

	NKK level	Text	Series
1	Harakeke e	He Kuri	He Purapura
2	Harakeke i	Ngā Manu i Runga i te Rākau	He Purapura
3	Kiekie a	Taniwha taniwha	He Purapura
4	Kiekie e	Poaka Kunekune	He Purapura
5	Kiekie i	Hōhepa te Pūru	He Purapura
6	Pīngao a	He Kai mā te Ika	He Purapura
7	Pīngao e	Kūri me te Ngaro	He Purapura
8	Pīngao i	Tamaiti Koioio	He Purapura
9	Pīngao o	Raraina Tuna	Ngā Kete Kōrero
10	Miro	Te Mocomoko	Ngā Kete Kōrero

Ranking these texts using method M2 showed a strong correlation with the ranking method used by Benton et al. (1995). The Spearman Rho rank order correlation coefficient for these two methods was 0.93 which was statistically significant ($p < .01$).

5.6 Ranking more advanced texts using *RANGE*

Following on from this preliminary set of texts, 33 higher level reading texts were analysed through *RANGE* and *WordSmith* to generate a pool of texts from which to select the series to be ranked by students and teachers. The data for this larger pool of texts is displayed in Appendix 5.

5.7 Selection of the text series for this study

Two series of six texts were selected for students and teachers to rank. Six texts were included in each series because it was considered that the task of reading and ranking any more than six texts at one sitting would be too onerous and time consuming for teachers and students. The series were selected using different methods, M1 and M2 (as described earlier). Where possible, texts were kept as neutral as possible which meant avoiding texts that had iwi-specific content knowledge. The texts that were eventually used in the two series were written before 1998. This was an attempt to find texts that had a greater chance of being unseen by the students. It was earlier noted by the researcher that most of the texts that students in year 5 to 8 classrooms had not already read were from the older collections of material.

5.7.1 Blue series

In selecting a group of texts using the vocabulary index (M1), the texts needed to have enough 'distance' between the vocabulary indices to make it more likely that participants would notice a difference between the texts to assist them to rank them successfully. Easiest is ranked 1, hardest is 6. Column 5 in Table 5.7 shows the differences between the Vocabulary Index score of each of these texts.

Table 5.7 Blue series ranked by Vocabulary index (M1)

Title	Number of content types	Raw vocab score	M1 Vocabulary Index	M1 differences between texts:
1: He Matakite Taku Hoa	202	603	2.99	
2: He Mahi Tinihanga	73	245	3.36	0.37
3: A Pāpaka Rāua ko Koura	65	229	3.52	0.26
4: Mōkai tuna	224	853	3.81	0.34
5: Te Tangihanga	180	725	4.03	0.24
6: Taringa Hōkeke	86	404	4.70	0.67

It was aimed to have at least a .20 difference between the Vocabulary Indices of each text. Text 5 and 6 had an even greater difference, with text 6 also being a much shorter text than text 5. The rationale behind this was to present a range of text length in relation to difficulty.

5.7.2 Orange series

The second series was selected by multiplying the Vocabulary Index by the number of function types (M2). Once again the texts needed to be sufficiently far apart using this method to assist participants to notice a difference in the texts so they could successfully rank them. Easiest is ranked 1, hardest is 6. Table 5.8 shows the set of texts selected as the Orange series.

Table 5.8: Orange series ranked by vocabulary index x function types (M2)

Title	Number of content types	Raw vocabulary score	Vocabulary Index	Function types (list 10)	(M2) Vocab Index x function	M2 Differences between texts:
1: Waimarie He Moemoea Noa Iho	42	156	3.71	38	141	
2: Te Kēhua o Waimā	63	272	4.32	40	173	32
3: Te Ana o Te Rau	142	488	3.44	60	206	33
4: Kia Tūpato	121	431	3.56	71	253	47
5: Te Nawe a Ngā Rākau	182	727	3.99	76	304	36
6: Kurī Heahea	194	723	3.73	88	328	21

Intercorrelation tables are presented in the next chapter for these two methods showing how well the two series compare with each other. The series of texts that were used in the validation studies of Elley’s noun count method were found to have different text features or author styles which occasionally produced lower correlations with their criterion measures. It is to be expected that this will also be the case for these two series, especially since they have been selected using different methods. All of the texts used in this study are referenced in Appendix 6 and the *RANGE* data is in appendix 7.

5.8 Summary

Phase Three has now brought the study to the position of having two series of texts selected and ranked. *RANGE* has proven to be a very useful tool for comprehensive analysis of the vocabulary used in texts, and has enabled a vocabulary index and a measure of syntactic density to be measured relatively

easily. The good correlation shown in the results of the comparisons with Benton's et al. earlier methods is encouraging. The 33 texts that were analysed showed between 81-96% coverage provided by just over the first thousand words with an average coverage of 92%. This is lower than the threshold recommended by Nation and is worthy of further investigation to see what the percentage of coverage is achieved across a much larger body of texts.

Because it would be possible to apply a variety of methods to the data, it was decided to explore some other common methods of measuring text difficulty at this point. The following chapter presents the additional methods that were tested.

Chapter Six: Additional methods for estimating text difficulty

In the previous chapter M1 was explained as the selection criterion for the Blue series of texts and M2 for the Orange series. Both of these methods use word frequency measures in some form or another. In this chapter, a comparison of the Blue and Orange series using M1 and M2 methods is undertaken to see if these methods produce the same rank order for both series of texts. Five other methods were also tested, and this chapter outlines how each of these methods were formulated and presents a comparison of rank orders these methods produced across both series of texts.

6.1 Vocabulary Index: Method M1

This method which was used to select the Blue series of texts has previously been described in section 5.3. Table 6.1 shows how both series of texts compared when being measured by M1.

Table 6.1 Comparison of the Blue and Orange text series using M1 method

Text	Blue	Rank	Orange	Rank
1	2.99	1	3.71	3
2	3.36	2	4.32	6
3	3.52	3	3.44	1
4	3.81	4	3.56	2
5	4.03	5	3.99	5
6	4.70	6	3.73	4

Table 6.1 shows that the Orange series does not follow the rank order of difficulty that the Blue series does with the application of the M1 method. Furthermore, there is much smaller differences between the texts of the Orange series with a range of .88 from text 1 through to text 6, while the Blue series produced a range

of 1.71. It was therefore not expected that the criterion measures would validate this method for the Orange series. See for example, Orange texts 1 and 6, where there is only a .02 point difference between them in their rank position as 3 and 4. This shows that the rankings made using method M1 do not generate comparable rankings across these text series. It also shows that validation using the criterion measures needs to be presented separately for both series.

6.2 Vocabulary Index x Number of function types: Method M2

This method which was used to select the Orange series of texts has previously been described in section 5.4. Table 6.2 shows how both series of texts compared when being measured by M2.

Table 6.2 Comparison of the Blue and Orange text series using M2 method

Text	Blue	Rank	Orange	Rank
1	307	6	141	1
2	181	1	173	2
3	187	2	206	3
4	305	5	253	4
5	294	4	304	5
6	291	3	328	6

Table 6.2 shows that using the M2 method, text 1 of the Blue set which was ranked easiest using the M1 method, now ranks as the most difficult. There is also only a 3 point difference between rankings of 3 and 4 for the Blue series. The range of difficulty from text 1 to text 6 for the Orange series using M2 is 187, while the range for the Blue series is much smaller at 126. As for M1, it is shown that the criterion used to select and rank texts by M2, also does not produce the same rank result across both series, and that the distance in measures between texts alters significantly.

6.3 Number of function word types: Method M3

The texts were also ranked by a method of counting the total number of function word types alone, which is a purely syntactic measure. Benton et al. (1995) suggested that a quick estimate of the linguistic difficulty of a Māori text could be gauged by counting the number of different function words that occur in the text. Benton et al. provide two lists: list one contains particles, prepositions, and directional adverbs; list two contains pronouns, demonstratives, possessives, conjunctions, and articles. As mentioned in 4.3.12, these two lists were the basis upon which the function word list (base word list 10) for this study was constructed. Using *RANGE*, the number of function word types is automatically produced in the analysis, being all of the words grouped from list 10. Table 6.3 shows how the two series ranked using M3.

Table 6.3 Comparison of the Blue and Orange text series using M3 method

Text	Blue	Rank	Orange	Rank
1	103	6	38	1
2	54	2	40	2
3	53	1	60	3
4	80	5	71	4
5	73	4	76	5
6	62	3	88	6

Using this method of measuring text difficulty, the Orange series retained its original rank order as when it was selected using M2. This was to be expected because the number of function types is a component of the M2 method. However, for the Blue series, the easiest text according to its original method of selection (M1), is now deemed to be the hardest because it has a high number of different function word types (103). Both series show a range of 50 points between the first and sixth ranked texts using M3.

6.4 Average sentence length: Method M4

Average sentence length is the second type of syntactic proxy considered for this study (labelled M4). Average sentence length was determined using *WordSmith* to calculate the total number of words (not characters) in a sentence. There are other readability measures that combine syntactic proxies such as sentence length with a semantic measure like word frequency. The *Lexile framework for reading* (1995) is an example of this, while others such as the Gunning Fog Index (1952) and Flesch Reading Ease Scale (1948) also incorporate average word and sentence length in their analysis of English texts.

Kamil (2001) says that sentence length is considered important by some, not only because it correlates with other measures, but directly, as a factor that affects reading in its own right. However, he also points out that this is still the subject of active research and debate. Pearson (1984) is cited in Kamil as having found that shortening sentence length does not automatically ensure easier reading or better comprehension, Elley (1982) found similar results when sentence length was shortened by simplifying complex sentences. Furthermore Spruck Wrigley (2001) says that ultimately, it is the relative complexity of the syntax, sometimes termed “heaviness,” along with the transparency or opacity of the sentence structures that causes difficulties for non-native speakers, not sentence length itself.

Klare (1963) theorised that sentence length was a factor in predicting text difficulty because of the load placed on verbal short term memory. This was supported by Crain & Shankweiler (1988) who argue that sentence length is a good proxy for the demands that syntactic complexity brings to a text. However, in the case of second language learners, the review on applying the *Lexical framework* which uses sentence length as a proxy claims that:

For those who are not fully proficient in the language they are trying to read, syntax often plays a much greater role than mere sentence length. Fairly short sentences that are easily understood by native speakers, passives for example, will present difficulties. (2001: 12)

Edwards (1999: 47), cites length of sentences and length of words as contributing to readability and provides the following as a guide for English texts:

Sentence length – Reading level (years)

About 7/8 words – 6 years

About 10/11 words – roughly 7/8 years

About 14/15 words – roughly 9/10 years

The average length of sentences (in words) for the two series selected for this study range from 6.8 to 17.2 words per sentence. Table 6.4 shows the average sentence length for each the texts and how the rankings fall using this method.

Table 6.4 Comparison of the Blue and Orange series using M4 method

Text	Blue	Rank	Orange	Rank
1	13.5	4	6.8	1
2	13.4	3	10.1	2
3	17.2	6	12.0	4
4	14.2	5	11.4	3
5	10.0	2	12.8	5
6	9.8	1	13.4	6

It is interesting to note in this comparison, that the average sentence length of the texts in the Orange series, follows the original M2 rankings very closely, while for the Blue series, sentence length does not align at all with its original M1 ranking. It could be concluded that the Blue series is less ‘internally stable’ meaning that various features of text difficulty do not match up within the texts. For example, simple high frequency vocabulary might be used, but with a very complex

syntactic style, or very long sentences. Conversely there may be very difficult vocabulary written into short simple sentence structures. For the Orange series it would appear that the length of sentence increases fairly evenly with an increase in the number of function words used, and the difficulty of the vocabulary used in the text. The range produced a difference of 7.4 for the Blue series and 6.6 for the Orange series.

6.5 Mean segmental type:token ratio: Method M5

Type:token ratio (TTR) is a measure of the number of different types of words in a text compared to the total number of tokens or words counted overall. Maxwell & Benton (1994) incorporated type token ratio into their method for ranking texts. The higher the ratio of types to tokens, the greater the lexical richness, and therefore the more difficult a text is presumed to be. When words occur more often in a text, the reader has several chances to tackle a word they may not know. The repetition factor greatly increases reading ease and gives the reader more opportunity to check their prediction of meaning in a variety of grammatical contexts. Laufer & Nation (1995) point out that lexical variety as shown by TTR is dependent upon the definition of a word. If derivatives are counted as different words, the variety will appear greater than if words are grouped into families. It is important to make it clear that words for this study have not been grouped into families in the type:token calculations. The other aspect that the TTR measure does not provide, is the frequency level of the words being used. Baayen (2001) cautions that type:token ratios are not independent of sample size and that in natural vocabulary, the type:token ratio will increase with text length. Richards & Malvern (1997) say that TTR is flawed unless calculated by standardising the number of tokens. This aspect was taken into consideration when calculating the type:token ratio for the texts used in this study. The default setting in *WordSmith* standardises the type:token ratio across groups of 1000 words. *WordSmith* was reset to calculate the standardised type: token ratio across 50 words. This means

that rather than a type:token being calculated across the whole text, it is calculated across the preset number of every 50 running words, then averaged across the whole text. This gives a more precise measure when comparing texts of differing lengths. This approach was considered necessary for this study to even out the effect for texts that were only 200 words in total compared to those in excess of 1300 words. Richards and Malvern describe this as Mean Segmental TTR (MSTTR), and here it is referred to as the M5 method. Changing from TTR to MSTTR, produced a different rank order for the texts in both series. Table 6.5 shows how the two series ranked using this method.

Table 6.5 Comparison of the Blue and Orange text series using M5 method

Text	Blue	Rank	Orange	Rank
1	73	5	62	1=
2	68	2	62	1=
3	69	3	65	3
4	71	4	70	4
5	66	1	72	6
6	74	6	71	5

The rankings produced using this method showed that both series retained a reasonably steady incremental rank order with the exception of texts 1 and 5 in the Blue series which switched rank. The Orange series had a rank change between texts 5 and 6, and a tie for texts 1 and 2. The range for the Blue series was 8 and for the Orange was 10. Although this is a small sample size, it indicates that this method might be fairly reliable regardless of other features that may be present in the text because it compares reasonably well across both series.

6.6 Vocabulary Index + average sentence length: Method M6

As mentioned previously, the readability method used in *Lexile Framework for Reading*, has adopted sentence length as its index of syntactic complexity which it combines with a semantic measure using word frequency. Adams says that:

Moreover, an “on average” relationship between sentence length and syntactic complexity is logically compelling: The longer the sentence, the greater the number of concepts in reference; the greater the number of concepts in reference, the greater the potential number or complexity of the interrelations that must be understood between and among them. (2001: 20)

Hancioglu and Eldridge (2007: 32) found that readers’ perception of difficulty and the average number of words per sentence was quite high, and they suggested that the raw average number of words per sentence may be a good indicator in an L2 context of likely text difficulty. Furthermore they concluded that including both lexical and structural measures in tandem will give a broad overview of text difficulty. Elley & Croft (1989) do not appear to have tested the noun frequency count method in combination with average sentence length. It was therefore decided to combine these two measures for this study and test for any validation with the criterion measures. Table 6.6 shows the comparison of using this method across the series.

Table 6.6 Comparison of the Blue and Orange text series using M6 method

Text	Blue	Rank	Orange	Rank
1	16.49	3	10.5	1
2	16.76	4	14.4	2
3	20.72	6	15.4	4
4	18.01	5	15.0	3
5	14.03	1	16.8	5
6	14.50	2	17.1	6

The Orange series of texts has remained reasonably true to its original rank, again showing the stability of this series of texts no matter the measure applied. The Blue series shows a very mixed rank order from its original ranking method. The range produced in these rankings was 6.69 for the Blue series and 6.6 for the Orange series.

6.7 MSTTR + number of function types: Method M7

The most commonly applied readability formulae use one semantic proxy and one syntactic proxy. One method of generating a semantic proxy is to produce a vocabulary index based on word frequency counts (already tested in this study as M1). Another, which has also been tested as M5, is to calculate a type:token ratio. This gives an estimate of lexical burden by measuring the 'richness' of a text. Generally, type:token ratio is produced using all of the words in the text, including function words. However, this means the type:token ratio is not producing a measure of strictly lexical burden. An attempt was made to produce an MSTTR of only the content words. This was tried by using a stoplist of function words for *WordSmith* to exclude from its type:token calculations. The number of different function types was then added to this 'content only' type:token ratio to provide a syntactic proxy. Unfortunately, it was discovered later, that *WordSmith4* was not excluding the stoplist of words in the statistics it produced for the type:token ratio. Mike Scott (personal communication), is currently working on a solution to this for *WordSmith5*. Notwithstanding this, comparisons which had already been undertaken, showed that this method which was now understood to be the MSTTR (for all words in the text), added to by the number of different function words, corresponded very strongly with methods M2, M3 and M5. This being the case, this accident in design has been included as a method for consideration for three reasons: it is much simpler to calculate than M2, it gives a more robust overview of the whole text than M3 because it is

calculated using all words, and its strength with M5 showed the strongest agreement across both series.

Table 6.7 shows how the texts in both series rank when this method is applied.

Table 6.7 Comparison of the Blue and Orange text series using M7 method

Text	Blue	Rank	Orange	Rank
1	176	6	99	1
2	122	1=	102	2
3	121	1=	126	3
4	151	5	140	4
5	140	4	147	5
6	136	3	159	6

Table 6.7 shows that this method produces the same rank order for the Orange series as its original selection method of M2. The Blue series shows a mixed rank order from the original M1. Note that text 1 (easiest) in the Blue series original rank has now become the hardest in rank order at 6. This also occurred when M3 and M2 were applied to the Blue series. The range for the Blue series was 55 and 60 for the Orange series.

6.8 Intercorrelation of methods

In the following two tables, a spearman rho rank order correlation coefficient was used to show how well each of the methods were related to each other, or will produce similar rankings across the separate series of texts. The critical level of $\rho = 0.89$. It has already been shown that only M5 produces similar results across both series, but it is also worthwhile to see which of the methods agree with others when applied to the same series. Table 6.7 shows the correlations between the different methods used to rank the Blue series of texts.

Table 6.8 Blue series: Intercorrelation matrix of Methods M1 – M7

	M1 Rank	M2 Rank	M3 Rank	M4 Rank	M5 Rank	M6 Rank
M2 Rank	-.09					
M3 Rank	-.14	.94*				
M4 Rank	-.54	.09	-.09			
M5 Rank	.09	.37	.31	-.09		
M6 Rank	-.43	-.26	-.37	.89*	.03	
M7 Rank	-.14	.94*	1.00*	-.09	.31	-.37

The critical r ($n=7$) = 0.89, $p < .05$

The methods showing statistically significant correlations with each other for the Blue series, are M7 with M2 (.94) and M3 (1.00). M7 and M2 are both a combination of semantic and syntactic proxies which may explain their strength of relationship. In addition, M3 shows strength with M2 (.94). M3 shares the element of function words with M2, which probably explains the strength it showed with M2. The strength that M4 showed with M6 (.89) was also to be expected because they both contain a measure of average sentence length. M1 (vocabulary index) is the method that this series of texts was selected by. The correlations show that M1 showed no strength with any of the other methods used.

Table 6.8 shows the correlations between the different methods used to rank the Orange series of texts.

Table 6.8 Orange series: Intercorrelation matrix of Methods M1 – M6

	M1 Rank	M2 Rank	M3 Rank	M4 Rank	M5 Rank	M6 Rank
M2 Rank	.09					
M3 Rank	.09	1.00*				
M4 Rank	.03	.94*	.94*			
M5 Rank	.13	.92*	.92*	.85		
M6 Rank	.03	.94*	.94*	1.00*	.85	
M7 Rank	.07	1.00	1.00	.94*	.93*	.94*

The critical r ($n=7$) = 0.89, $p < .05$

The Orange series also showed strong correlations for M7 with methods M2 and M3 and in addition, with M5. This could be a result of these four methods using every word type in the measurement method and therefore representing both semantic and syntactic proxies. These four methods have all shown statistical significance with all other methods except for M1 which is the vocabulary index alone. The strong association of M4 with M6 was also shown again for this series (1.00). Overall, M2, M3, M5, and M7 show that they will produce similar results to any of the other methods except for M1. M1 has been shown across both series of texts to have no strength of correlation with the other methods used.

6.8 Summary

This chapter has outlined five further methods of measuring text difficulty and has also shown that each of the texts in the two series occupy different rank orders each time a different method is applied. The main finding to emerge from exploring these methods as a group, was that the Blue series of texts showed itself to be less stable across differing measurements of text difficulty. In contrast the Orange series maintained a reasonably consistent rank order when various methods of measuring text difficulty were applied, with the exception of M1. This suggests that the Blue series which was selected on vocabulary load alone, was

relatively unstable. The other main finding of this section was that the only method that showed it could produce a similar rank order to the original method of selection across the two series was M5, mean standardized type:token ratio. Method M7 showed itself to correlate very closely with M2 M3 and M5 which, with the added benefits of ease of calculation and the inclusion of all words in the text would make M7 the method of preference.

The process of applying various methods to measure difficulty of texts has shown the wide range of challenge that can be built in to the linguistic components of a text. Writers can use simple vocabulary, but then write with very long sentences, or use a very complex syntactic style. Conversely, a text with an even spread of difficulty across features of text will have vocabulary, sentence length, syntactic complexity, and lexical richness all tied together in a close range of difficulty. By revealing a global view of the challenge the components of a text may pose, this part of the study has shown that an analysis of the complete 'build' of a text has shown how well rounded and balanced aspects of text difficulty need to be. This suggests that valuable feedback can be provided to authors about a range of important aspects for texts that are specifically written to meet levelling criteria using a controlled vocabulary approach. This has implications for monitoring the make up of the text at the early stages of levelling, should this be desired.

Chapter Seven: Phase Four Establishing Criterion Measures of Text Difficulty

Following on from phase three, the six methods used to rank the texts were now ready for validating against criterion measures. Phase four was the establishment of the criterion measures to be used and also the identification of the strongest criterion measure. The criterion measures developed in this phase are teacher rankings, student rankings, combined rankings, and student performance. In addition to this, a questionnaire was completed by teachers. This is reported on separately in Chapter 9. While establishing the criterion measures, the participant groups of teachers and students had no knowledge of the rankings, or the methods by which the texts had been ranked. The procedures for collecting reader opinion and student performance to establish the criterion measures are described in this chapter.

7.1 Teacher and student opinion

Teacher and student opinions and student performance were sought in order to validate the methods used in this study to estimate readability. Elley & Croft say in relation to the criterion measures they used that:

Teachers' and pupils' opinions were used as a criterion measure because they provided a more direct and sensitive measure than the usual method of graded texts or results from comprehension tests. Studies conducted by Klare (1974) and Harrison (1980) have confirmed that pooled teacher judgements are an excellent criterion for judging the validity of various readability estimates. (1989: 15)

The following sections detail how teacher and student opinions and student performance rankings were obtained as criterion measures.

7.2 Participants

The student participants involved in this part of the project were from a kura kaupapa Māori and ranged from year 8 to year 13. There were ten students involved in the ranking task, eight of whom also undertook the performance tasks. All of the students are second language learners of Māori.

The 15 teacher participants comprised 2 teachers from the field site kura, 6 resource teachers of Māori, 2 university lecturers, and 5 teachers involved in a bilingual postgraduate training programme. Of this total group three are first language speakers of Māori. The group was drawn from throughout the North and South Islands of New Zealand and represents a range of iwi affiliations. The teachers were not asked to undertake the performance task but were asked to reflect on the texts using a questionnaire as they ranked each text. All participants were coded as shown in Table 7.1 to preserve anonymity.

Table 7.1: Examples of participant codes

Code	Participant
TO1	Teacher one who ranked the Orange series of texts.
SO1	Student one who ranked the Orange series of texts.
TB4	Teacher four who ranked the Blue series of texts.
SB5	Student five who ranked the Blue series of texts.

7.3 Ranking task

Because it was considered important to keep the rankings as individually constructed as possible, the student ranking of the texts was administered as a classroom activity under exam-type conditions, supervised by the researcher. The performance task was administered individually by the researcher, no later than a week after the ranking task. For the Blue series, five students completed

the ranking task and three did the performance task. Five students completed both tasks for the Orange series. It took most students one hour to read and rank the texts and a further 40 minutes to complete the performance task.

The texts for the teacher group were posted to individuals and then returned to the researcher. Seven teachers ranked the Blue series of texts and eight teachers ranked the Orange series.

All participants were asked to rank the six texts from 1 to 6, 1 being easiest and 6 being hardest. No suggestion was made to either group as to what constituted difficulty so that opinions were not influenced in any way by the researcher. The texts were given to them as a complete series placed in a random order. The ranking sheets are in Appendix 9. Both groups were asked to highlight unknown words, and to mark any reading re-runs they did. For example, if they returned to the beginning of a sentence, word, or any place in the text as a strategy for gaining meaning. This was done to draw their attention to the difficulties each text presented them with. As they reviewed the series in order to rank the texts, the markings they made would help them to remember the challenges each text presented. The teacher group was asked to consider their rankings on a professional level as if they were selecting texts for children to read, rather than difficulty from a personal perspective.

Maxwell and Benton (1995) carried out a similar ranking task, but with a smaller group of six year 9 students and one teacher, using a series of seven texts. The method used for their computed ranking was a combination of type:token ratio, occurrence of words at various levels, unusual words, and the total number of words. Maxwell and Benton reported 'considerable agreement, especially between the computer and the students' (1995: 3). Using Maxwell & Benton's data which were reported in the vocabulary levels section Spearman Rho correlation coefficient analyses were calculated by the researcher.

Table 7.2 Spearman Rho correlation matrix from Maxwell and Benton's (1995) ranking task

	Student ranking	Teacher ranking
Computer ranking	0.54	0.46
Student ranking		0.79

This analysis shows that the teacher ranking and student rankings showed the highest correlations with each other (0.79), but neither teacher ranking (0.46) nor student ranking (0.54) showed a strong relationship with the computer ranking.

7.4 Student performance tasks

Hancioglu and Eldrige (2007) say that while reader intuitions are certainly valuable, they also need to be treated with caution and, if possible, matched against more objective data. The cloze test method (Bormuth, 1966) was a criterion measure used to validate the results in Elley's research. Supporters of the cloze procedure say that it provides a solid measure of both semantic and syntactic challenge. Anderson asserts the validity of this technique for second language learners in saying that:

The number of words a subject replaces correctly is an index of his comprehension of the passage, and of the reading difficulty of the passage for that reader. Cloze procedure is suitable as a measure of reading comprehension for non-native readers of the language. (1971: 178).

Moyle (1970) maintains the benefits of using the cloze procedure are that it includes accuracy of vocabulary, fluency, knowledge of grammatical structure and understanding of the text and, therefore, is a measure of total readability.

In contrast to this view, however, Stephens cites the following criticisms of the cloze procedure:

In particular, critics suggest that cloze is inappropriate for measuring text or readers' abilities in languages other than their native language. The results of cloze testing reflect the reader's basic intuition about the structure and vocabulary of the target language -- and that does not exist for the language student (2000: 1).

Cloze testing alone was not considered to be sufficiently accepted in the second language learner setting to be used as the primary measure of comprehension. It has, therefore, been incorporated into this study with caution. Some of the multi-choice questions used a cloze style of question. For this study multi-choice and free explanation were considered to be the most manageable and time effective comprehension tasks to administer.

7.4.1 Constructing the student performance tasks

Eight multi-choice questions were written for each text. There were four choices of answer provided for each question. Three Resource Teachers of Māori independently provided feedback to the researcher about the clarity and quality of the multi-choice questions and the answer choices provided for each question. Most of the questions were kept at a level of what Herber (1978) describes as level one, requiring students to locate or recall information in a text; and level two, to interpret what the author means. They each contained a question which sought to identify the writer's voice, for example first person, or third person, and some followed a cloze or completion style of question. Care was also taken to ensure that minimal new, and no difficult vocabulary was introduced into the questions or choices. Questions didn't seek to clarify word meaning as this was covered specifically in the vocabulary check. The words to be tested in the vocabulary check were drawn from the *RANGE* analysis of words identified as

not being on the lists, or from the higher end lists (see Appendix 7). There were generally 10 words selected.

7.4.2 Administering the student performance task

At a time no more than one week following the initial reading and ranking task, students were given 8 multi-choice questions per text to complete individually. They also had access to their original set of texts. Time was left between these tasks; firstly to fit with kura time frames and what was reasonable for time spent on a task of this nature, and secondly, the time delay meant that students usually chose to re-read or at least scan the text for a second time. This was consistent with the rationale for providing a second exposure as discussed in section 3.3.4. The researcher was present during this time to ensure there was no collaboration between students on this task.

As students completed the multi-choice section, they were individually tested on their knowledge of the unusual or critical content vocabulary used in each text. This was purposefully done following the multi-choice task because engaging closely with these lesser known words during the vocabulary check had the potential to influence their performance on the multi-choice task. For the vocabulary check, students were asked to supply either a synonym, or free explanation of the meaning of the word, which was highlighted and read in context for them from the text. The researcher recorded responses. The results of the multi-choice task and vocabulary check were then combined to produce the student performance scores. The performance tasks are contained in Appendix 8.

7.5 Criterion measures

The procedures carried out in this phase produced four criterion measures, teacher opinion, student opinion, combined opinion and student performance. The teacher and student rankings were averaged and ranked again to produce the combined rank. The student performance scores were also averaged for each text, and the texts ranked again using the averages. These rankings were then analysed using Spearman's Rho corrected for ties, to identify strength of correlation between the criterion measures and a range of methods for estimating readability.

7.6 Intercorrelation of criterion measures

Once this phase had established the criterion measures, they were intercorrelated using Spearman's Rho to identify the strength of correlation between the measures.

The ranked opinions of teachers and students have been kept separate to make more transparent the role that each group has played in producing these results. For example, in some of the methods tested, the correlations with the teacher opinion was very strong while the student opinion barely rated and vice versa. This caused the correlations and statistical data of the combined opinion to be inflated by one group. It was important to be able to see the spread and balance of opinion between teachers and students or it may have been that some of the real opinion became buried within the opinion of another group. If this was not brought forward and made clear, the results might have been representing only one dominant group throughout the testing of all methods. This is why the averaged rankings of teacher and student opinion are presented as separate criterion measures.

Table 7.3 shows the intercorrelation matrix of criterion measures for the Blue series and Table 7.4 for the Orange series. Table 7.5 shows the median of those criterion measures across both series.

Table 7.3 Intercorrelation matrix of criterion measures for Blue series using Spearman’s Rho

	Student rank	Combined rank	Student performance
Teacher rank	0.57	0.99*	0.50
Student rank		0.69	0.60
Combined rank			0.60

*p<.05

Table 7.4 Intercorrelation matrix of criterion measures for Orange series using Spearman’s Rho

	Student rank	Combined rank	Student performance
Teacher rank	0.70	0.77	0.64
Student rank		0.99*	0.74
Combined rank			0.75

*p<.05

Table 7.5 Median of intercorrelation coefficients between criterion measures for Blue and Orange series using Spearman’s Rho

	Student rank	Combined rank	Student performance
Teacher rank	0.64	0.88	0.57
Student rank		0.84	0.67
Combined rank			0.68

*p<.05

The intercorrelation matrices show that within the criterion measure of the combined rank, the teacher opinion inflated the results for the Blue set, (0.99 between teacher and combined rankings) while for the Orange set the student opinion dominated (0.99 between student and combined rankings). Table 7.5 shows that overall across both series, the combined rank is the strongest of the criterion measures, almost reaching significance at 0.88. The critical level of rho =0.89.

7.8 Summary

This section has described the process of collecting data to produce criterion measures for testing the validity of the proposed methods of measuring text difficulty. Once all of the criterion measures were intercorrelated, the combined rank was shown to have the strongest intercorrelations for both series of texts reaching 0.99 and showed a median of 0.88 across both series. Elley & Croft (1989) reported correlations of over 0.90 for the combined opinions of teachers and students, which supported the use of combined opinion as a criterion measure for validating their studies. As shown in Table 7.2, Maxwell & Benton's (1995) findings also showed reasonable strength of correlation between student and teacher opinion (0.79). The findings for this study concur with that strength for both series, which showed the combined opinion would be a valid criterion measure to use. Having the criterion measures established, paves the way for Phase Five of the study which is the process of using the criterion measures to assess the validity of the seven methods applied to estimate the readability of the texts.

Chapter Eight: Phase Five

Validation of methods to estimate text difficulty

8.1 Introduction

This chapter presents the validation process of the vocabulary index (M1) and vocabulary index multiplied by function types (M2), along with the four other ways the trial texts have been measured for difficulty. The previous chapter described other methods, which were; the number of function types (M3), average sentence length (M4), mean segmental type:token ratio (M5), vocabulary index plus average sentence length (M6) and M7. The correlations of each method with the criterion measures is presented for both series of texts, Blue and Orange. It is recognised at this point that due to the small number of texts that were ranked it will be difficult to display results showing statistical significance.

8.3 Method of statistical analysis

Intercorrelation matrices were generated for both text series for all of the methods tested (M1-M7) using the Spearman Rho rank order method with the criterion measures (teacher rank, student rank, combined rank, and student performance). In some cases, rankings saw two texts tied for a place. As this impacts significantly when there are only a small number of texts, all of the rho coefficients reported have been corrected for ties. A $p < .05$ is used to indicate statistical significance. The critical level of $\rho = 0.89$.

In each of the following sections, each method of measuring text difficulty is examined in relation to the criterion measures. Note that in the graphs, the rank order of the texts changes, depending on the method applied to measure the texts. The criterion measure of combined rank does not appear on the graphs to

avoid clutter, but combined rank is reported in the tables. While the graphs give a visual presentation of the trends, it is the statistical evidence that carries the weight of evidence, and as such only the tables are discussed.

8.4 The vocabulary index method (M1)

This method of measuring difficulty has been described in section 5.3. and was the foundation method proposed for testing in this study. The order of text difficulty produced for the Blue and Orange series of texts are presented in the graphs with three of the criterion measures in Figures 8.1, and 8.2. The correlations are shown in Tables 8.1. and 8.2. It should be remembered that the Blue series of texts (see section 5.7.1) were selected using this method. Therefore, it was expected that if there were to be any strength of correlation shown for this method with the criterion measures, it would be stronger in the Blue series than the Orange series.

Figure 8.1 Blue series: The rank order of texts by difficulty for M1 and criterion measures

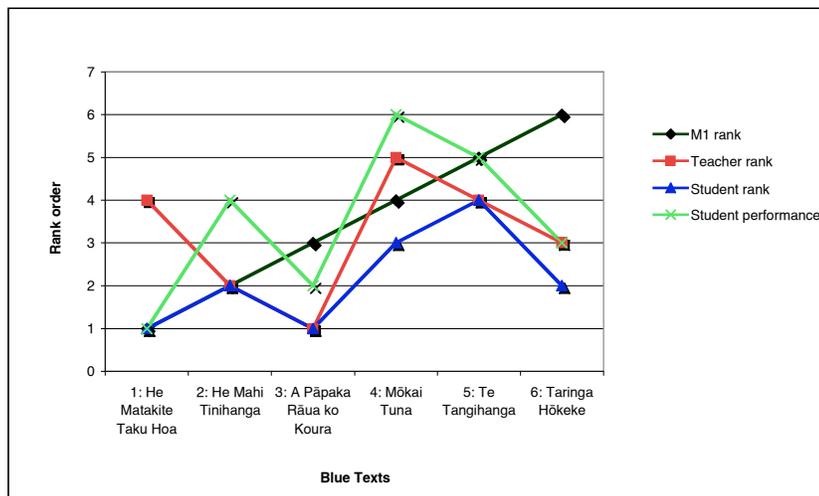


Table 8.1 Blue series: Correlations of M1 with criterion measures

M1 – Vocabulary index method	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	0.14	0.62	0.26	0.49

Table 8.1 shows that none of the rho’s are statistically significant using this method with the Blue series.

Figure 8.2 Orange series: The rank order of texts by difficulty for M1 and criterion measures

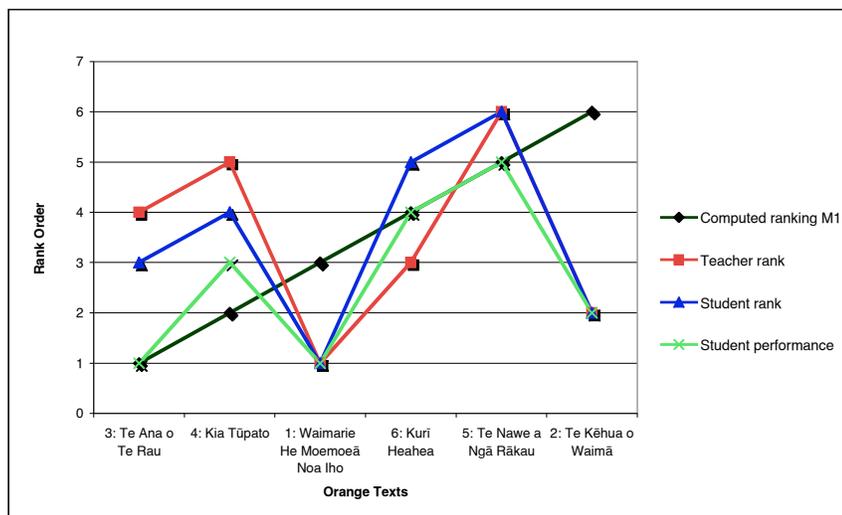


Table 8.2 Orange series: Correlations of M1 with criterion measures

M1 – Vocabulary Index Method	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	-0.14	0.20	0.09	0.49

Table 8.2 shows that none of the rho’s for the criterion measures show statistical significance using this method with the Orange series.

While the correlations with the criterion measures were slightly stronger for the Blue series than the Orange series, results for both series showed no statistical significance with criterion measures and measuring texts using this method. The

patterns in the graphs showed some agreement between all of the criterion measures.

These results may indicate that vocabulary load alone was not a sensitive enough measure for readers to distinguish differences between these texts. Correlations might be strengthened if the differences between measures in the texts was increased further. As mentioned previously, the small number of texts ranked made statistical significance difficult to show. A key hindrance in the word lists prepared for this study compared to those used by Elley & Croft (1989) is that the corpora used are from predominantly adult speech and writing. Elley and Croft (1989: 10) say that “the most useful lists were those derived from children’s own writing, rather than from books written by adults”. No studies have been undertaken to show that this would also be the case for Māori. In the opinion of Boyce, (private communication) it is not likely that there will be significant differences in the high frequency items in a corpus drawn together from Māori spoken *by* children and that written *for* children. A small pilot corpus of children’s spoken Māori has recently been analysed and early indications are that vocabulary production is consistent with the frequencies and coverage provided by the 10 baseword lists constructed for this study. However, Elley (1969: 421) says that “Comprehension appears to depend more on familiarity of words used by pupils in their writing rather than of the words they encounter in their reading”. Until further corpus development is undertaken which draws from the productive language of young speakers and writers of Māori, results will remain inconclusive about the reliability of the word lists used for this method based on Elley’s approach of measuring vocabulary load.

8.5 Vocabulary index multiplied by function types (M2)

Because the Orange series of texts was originally selected by this method, it was expected that if any strength of correlation were to be shown for this method with the criterion measures, it would be stronger in the Orange series than the Blue series. Figures 8.3 and 8.4 show the rankings graphed, and Tables 8.3 and 8.4 contain the correlation results.

Figure 8.3 Blue series: The rank order of texts by difficulty for M2 and criterion measures

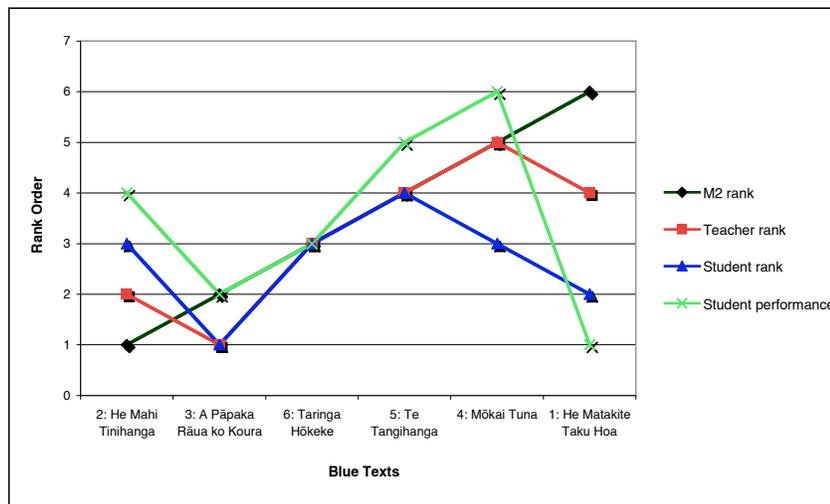


Table 8.3 Blue series: Correlations of M2 criterion measures

M2 – Vocabulary index x function types	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	0.84	0.09	0.78	-0.03

Only the teacher ranking showed support for using this method with the Blue series, just failing to reach significance at $p=.06$.

Figure 8.4 Orange series: The rank order of texts by difficulty for M2 and criterion measures

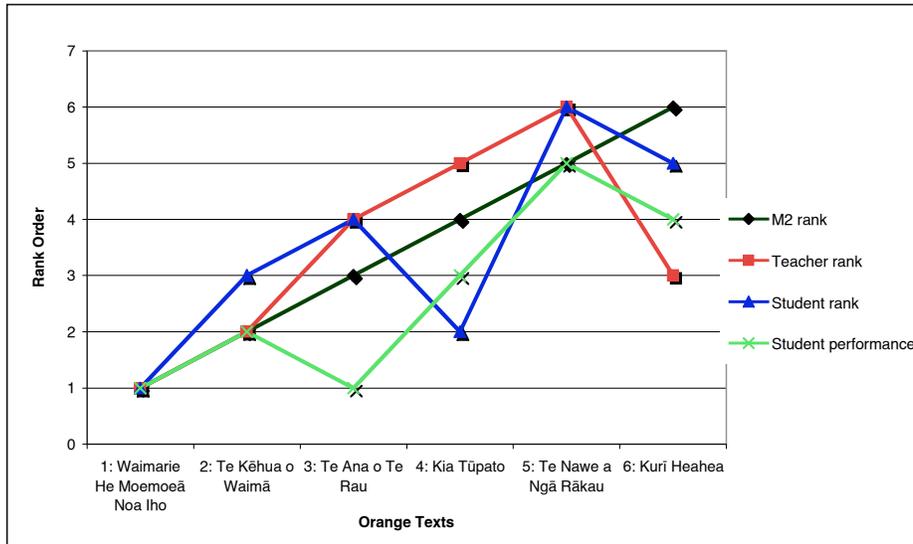


Table 8.4 Orange series: Correlations of M2 with criterion measures

M2 – Vocabulary index x function types	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	0.66	0.84	0.89*	0.84

*p<.05

Student ranking and student performance just failed to reach significance ($p=.06$), while the combined rank showed significance using this method with the Orange series (0.89).

There are many readability formulae that use a combination of vocabulary load and a syntactic proxy of some kind. The results presented above have supported that approach to some extent. As expected, the Orange series showed stronger correlations and statistical significance overall with this method. The Blue series shows strength with the teachers' ranking, but not with the students' ranking or performance. It is to be remembered that the Blue series was not selected using this criterion. For both series, the combined ranking results are promising. In the

combined rank, the Blue series came close to showing statistical significance, and the Orange series showed significance. These results show that the vocabulary index measure in combination with this particular syntactic proxy of function words, is showing slightly higher validity.

8.6 Number of function word types (M3)

The following Figures 8.5 and 8.6 show graphed results for the number of function types in the text, compared with criterion measures, followed by the correlation data in Tables 8.5 and 8.6.

Figure 8.5 Blue series: The rank order of texts by difficulty for M3 and criterion measures

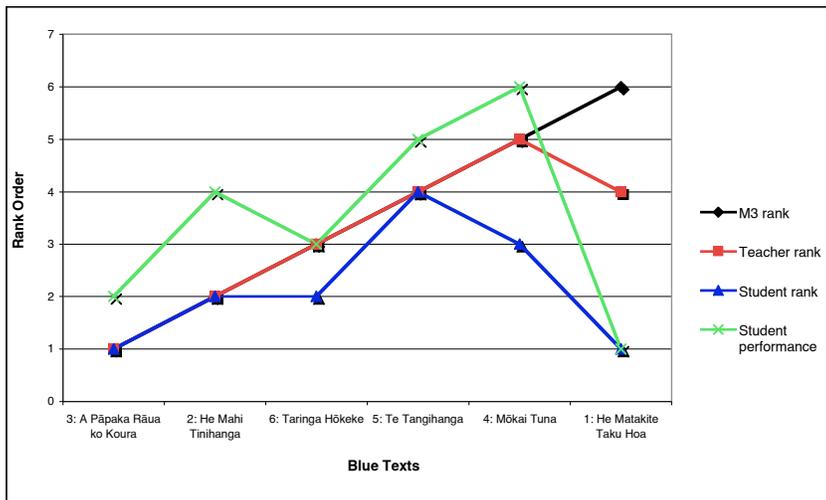


Table 8.5 Blue series: Correlations of M3 with criterion measures

M3 –Number of function types	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	0.90*	0.21	0.83	0.09

*p<.05

Only the teacher ranking showed any stastically significant agreement with this method for the Blue series, (0.90) and was higher than the combined ranking (0.83).

Figure 8.6 Orange series: The rank order of texts by difficulty for M3 and criterion measures

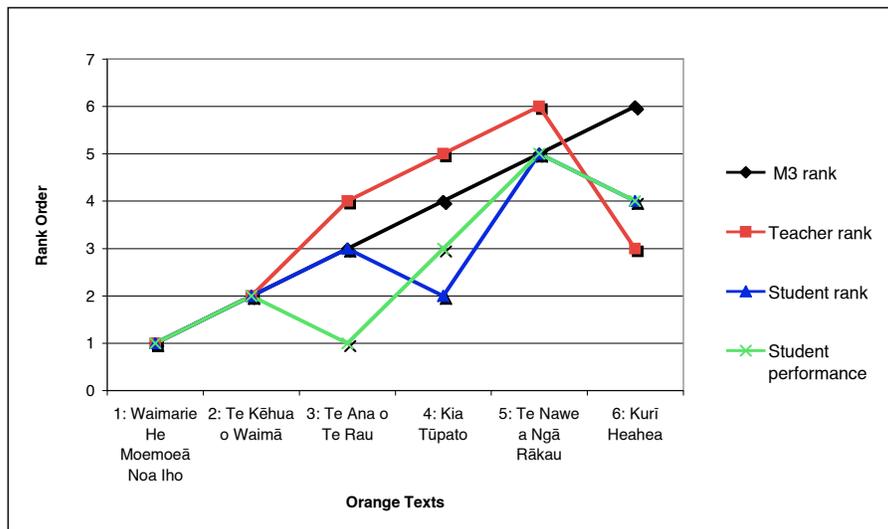


Table 8.6 Orange series: Correlations of M3 with criterion measures

M3 –Number of function types	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	0.66	0.84	0.89*	0.84

*p<.05

Table 8.6 shows much stronger support for this method, with a good spread of strength across all criterion measures, although only stastically significant with the combined ranking (0.89). Using M3, the Blue series of texts showed a strong correlation with teacher ranking (p<.05). However, this was not supported by student ranking or performance. In contrast, the Orange series of texts showed moderate to strong correlations across teacher and student rankings and student performance. These findings make this method worthy of further investigation.

8.7 Average sentence length (M4)

The following Figures 8.7 and 8.8 show the rankings (using average sentence length) graphed with the criterion measures, followed by the correlation data in Tables 8.7 and 8.8.

Figure 8.7 Blue series: The rank order of texts by difficulty for M4 and criterion measures

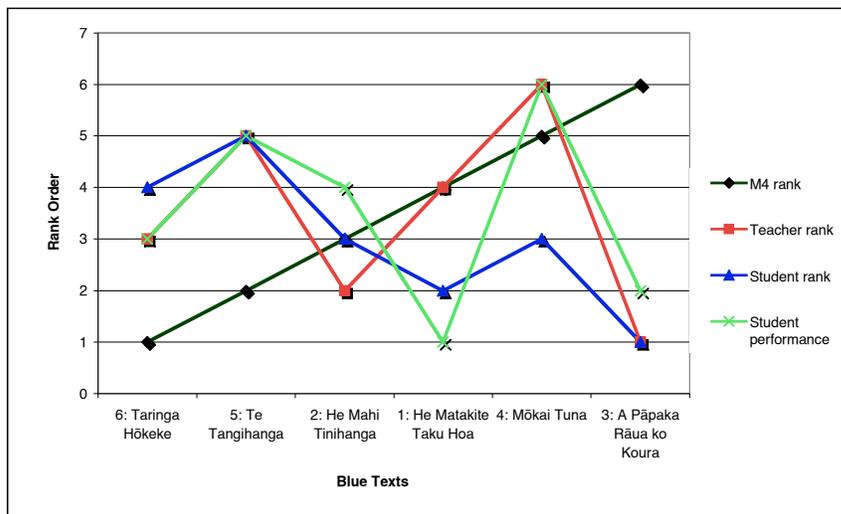


Table 8.7 Blue series: Correlations of M4 with criterion measures

M4 – average sentence length	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	-0.09	-0.44	-0.14	-0.14

The results in Table 8.7 showed no statistically significant relationships with the measure of average sentence length for the Blue series. In fact, the rho's were negative suggesting that texts with shorter sentence length tended to be more difficult than texts with longer sentences. This finding supports some of the claims made in some of the research discussed in section 6.4.

Figure 8.8 Orange series: The rank order of texts by difficulty for M4 and criterion measures

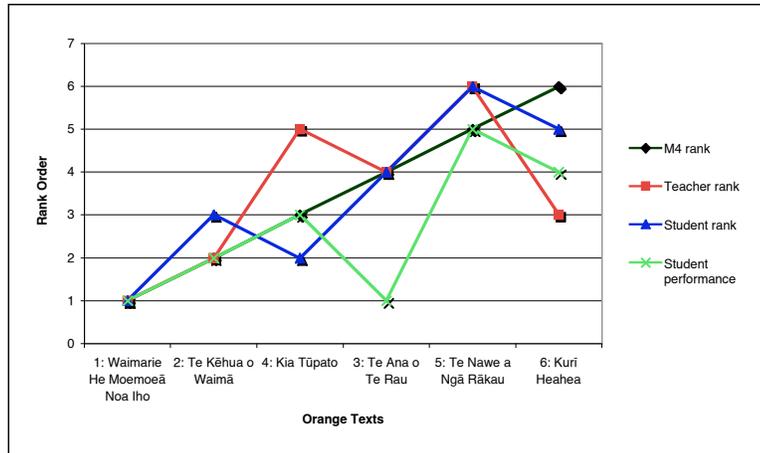


Table 8.8 Orange series: Correlations of M4 with criterion measures

M4 – average sentence length	Teacher ranking	Student ranking	Combined ranking	Student Performance
Rho	0.60	0.93*	0.94*	0.70

*p<.05

For the Orange series, it can be seen that the student ranking and the combined ranking showed statistical significance (0.93 and 0.94) and the strength of correlation with student performance was moderate (0.70). The teacher ranking alone showed moderate strength of relationship (0.60). Since this method generated some strength of correlation it would be worth including in further studies.

8.8 Mean segmental type/token ratio (M5)

Figures 8.9 and 8.10 show the rankings graphed with criterion measures for M5, and Tables 8.9 and 8.10 show the correlations for this method.

Figure 8.9 Blue series: The rank order of texts by difficulty for M5 and criterion measures

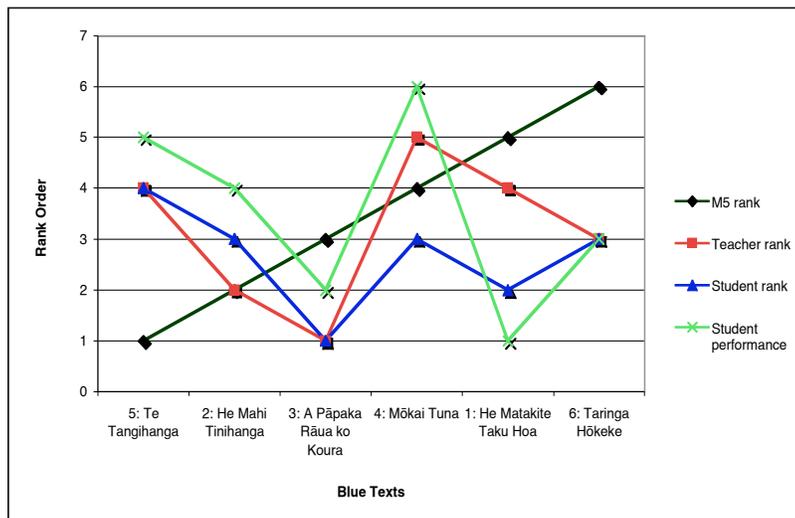


Table 8.9 Blue series: Correlations of M5 with criterion measures

M5 – mean segmental type token ratio (50)	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	0.14	-0.44	0.03	-0.43

No statistical significance was shown in these results, however, a negative relationship was shown. This suggests that a low standardised type:token ratio (a non-rich text), was associated with students' perception of difficulty and in the performance results, did not make comprehension easier for this series of texts.

Figure 8.10 Orange series: The rank order of texts by difficulty for M5 and criterion measures

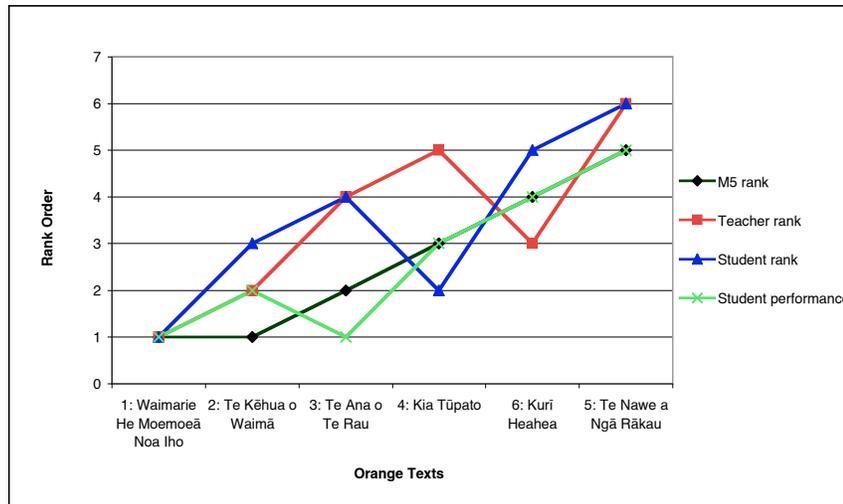


Table 8.10 Orange series: Correlations of M5 with criterion measures

M5 – mean segmental type token ratio (50)	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	.81	.87	.93*	.87

*p<.05

In sharp contrast, for the Orange series, this method has drawn the strongest agreement between any method and the criterion measures. The combined ranking (0.93) showed significance and both student ranking (0.87) and performance (0.87) just fell short of showing significance.

Within each of the series, the differences in the MSTTR between each text are greater in the Orange series than the Blue series. This means the Orange series gives the reader a slightly wider difference in richness between the texts, and was, therefore, more likely to draw agreement. Table 8.11. shows the difference in MSTTR between each of the texts.

Table 8.11 Range of MSTTR for both series of texts

Title: Blue series	MSTTR	Title: Orange series	MSTTR
5:Te Tangihanga	66.00	1:Waimarie he Moemoeā Noa Iho	62.00
2:He Mahi Tinihanga	68.00	2:Te Kēhua o Waimā	62.00
3:A Pāpaka rāua ko Kōura	68.50	3:Te Ana o Te Rau	65.25
4:Mōkai Tuna	70.89	4:Kia Tūpato	70.31
1:He Matakite Taku Hoa	72.77	6:Kurī Heahea	71.04
6:Taringa Hōkeke	73.75	5:Te Nawe a Ngā Rākau	72.44

The bigger differences in MSTTR between texts for the Orange series, excepting the tie between first two texts (which were often ranked the same by readers), could explain why the Orange series shows statistical significance with the criterion measures. Overall, this method has produced the best results so far, both with all criterion methods and rankings across both series.

8.9 Vocabulary Index + average sentence length (M6)

The rankings and criterion measures for this method are graphed in Figures 8.13 and 8.14 and the correlations are in Tables 8.12 and 8.13

Figure 8.13 Blue series: The rank order of texts by difficulty for M6 and criterion measures

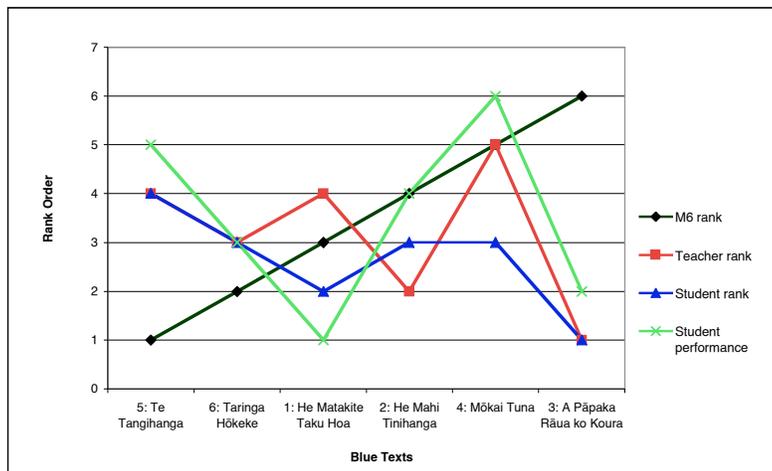


Table 8.12 Blue series: Correlations of M6 with criterion measures

M6 – Vocabulary index + average sentence length	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	-0.32	-0.48	-0.37	-0.09

The Blue series produced no statistically significant relationships with the criterion measures for this method, and the rho’s showed a negative relationship similar to that produced for the average sentence length.

Figure 8.14 Orange series: The rank order of texts by difficulty for M6 and criterion measures

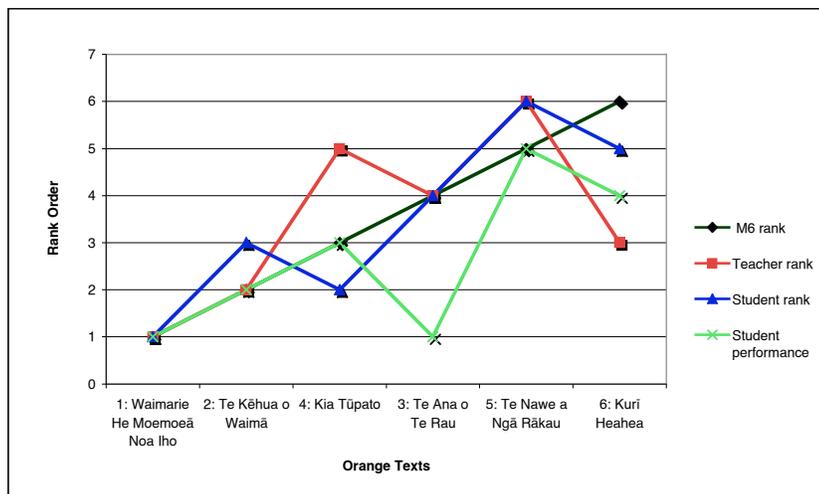


Table 8.13 Orange series: Correlations of M6 with criterion measures

M6 – Vocabulary index + average sentence length	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	0.60	0.93*	0.94*	0.70

*p<.05

The results shown here indicate that the Orange series produced significant correlations with the rankings from students (0.93) and in combination with teachers (0.94). There was moderate correlation with student performance (0.70) and teacher opinion (0.60).

8.10 MSTTR + function types

The rankings and criterion measures for this method are graphed in Figures 8.15 and 8.16 and the correlations are in Tables 8.14 and 8.15

Figure 8.15 Blue series: The rank order of texts by difficulty for M7 and criterion measures

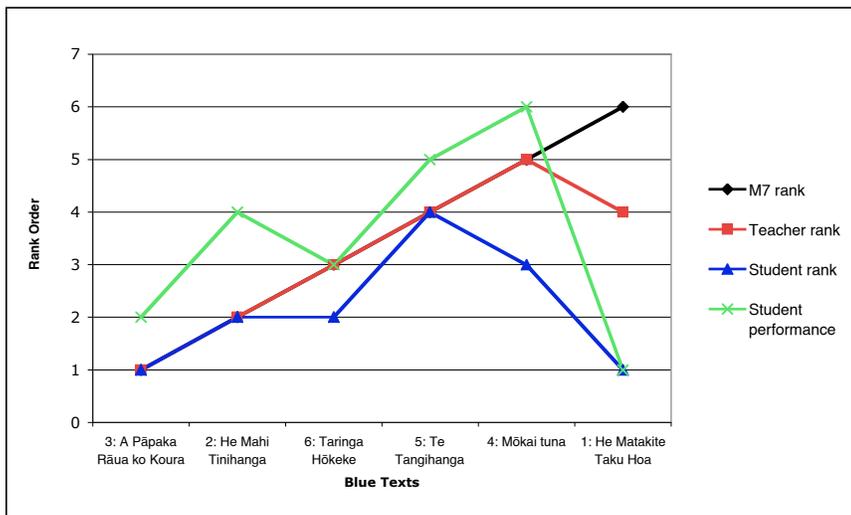


Table 8.14 Blue series: Correlations of M7 with criterion measures

M7 – MSTTR + number of function types	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	0.90*	0.21	0.83	0.09

* $p < .05$

These results show that this method produced significant correlations with the teacher rankings (0.90) and the combined ranking (0.83) was moderate to strong.

Figure 8.15 Orange series: The rank order of texts by difficulty for M7 and criterion measures

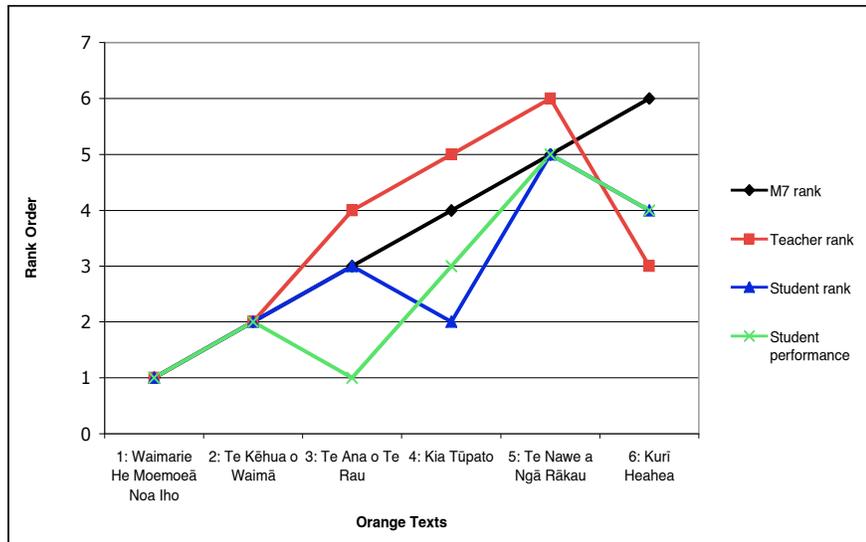


Table 8.15 Orange series: Correlations of M7 with criterion measures

M7 – MSTTR + number of function types	Teacher ranking	Student ranking	Combined ranking	Student Performance
rho	0.66	0.84	0.89*	0.84

*p<.05

These results show moderate to strong correlations across all criterion measures. It is promising to have the teacher ranking and student ranking contributing more evenly to the statistically significant result of 0.89. Student ranking and student performance were in total agreement, as they were for M5, the MSTTR method alone.

8.11 Overall validity of methods

The previous sections show the results of examining the criterion measures across both series using all six methods. In order to draw this information

together into a global result for the whole study, Table 8.14 shows the median rho's across criterion measures for both series in order to find the method with the strongest validity regardless of the series it was applied to. This way of displaying the overall picture is similar to the presentation of Elley & Croft's global findings for 12 methods using five different series.

Table 8.16 Rho coefficients between method and combined teacher and student rankings for both series

Method of estimating readability	Blue Series	Orange Series	Median
M1: Vocabulary Index	0.26	0.09	0.17
M2: Vocabulary index x function types	0.78	0.89	0.83
M3: Function types	0.83	0.89	0.86
M4: Average sentence length	-0.14	0.94	0.40
M5: Mean segmental type token ratio	0.03	0.93	0.45
M6: Vocabulary Index + average sentence length	-0.37	0.94	0.61
M7: Mean segmental type token ratio + function types	0.83	0.89	0.86

It can be seen that the Orange series produced strong correlations for all methods with the criterion measures with the exception of using the vocabulary index alone (M1). However, there is instability in the Blue series which produces negative effects for M4 and M6 which the Orange series shows as having strongest validity. It is difficult to explain why this inverse correlation has presented in the Blue series. M2, M3 and M7 are the only methods which can

show reasonable correlations with the criterion measures across both series with median rhos of 0.83, 0.86, and 0.86 respectively.

8.11 Summary

Throughout the correlation of methods with criterion measures across both series, the Blue series is consistently at odds with itself, while the Orange series maintains reasonable stability when measured by different methods. It is very tempting to draw conclusions from the results shown for the Orange series alone because of its stability factor, and give less credibility to the Blue series. However, because the reality is that in the world of children's texts in Māori, there is a differentiation between features of texts in their composition, we are challenged to accept the results from both series. It would appear from these results, that M2, M3 and M7 can be better relied upon, regardless of the make-up of the texts in a series being measured. However, this should not mean disregarding the results produced by M4, M5 and M6 in future studies. It is worthwhile noting that the addition of function types has strengthened the validity of the vocabulary index, because when standing alone as a measure (M1), it showed no validity at all. It was also shown in Phase One of this study, that function words did not show high self-correction rates for all groups (see Figure 3.4). This leads us to look more closely at the role of function words in determining text difficulty. It appears from these results, that including a measure of function words has been the critical factor in adding validity to the vocabulary index. Furthermore, M3 and M7 have shown equally in Table 8.16 that they have the strongest median correlations for both series. While M3 is calculated using only a syntactic measure, M7 is calculated using all words used in the text, and for that reason, it could be expected to give a more reliable measure overall.

Chapter Nine: Teachers' evaluations of the texts

9.1 Introduction

A questionnaire was designed to find out from the teachers what their overall response was to the text they had just read. This information was primarily gathered so that it took the teachers through a process of reflecting on the text in order to assist them later in the ranking process. It was believed that if they did a guided reflection of the same kind on each text, they would better align the texts when they ranked them. The teacher participants were given a questionnaire for each text. The questionnaire asked them to rate the grammatical complexity of the text using a scale from 1 to 4, whether there were many words unknown to them, how they dealt with unknown words, what strategies they used to make sense at the sentence level and the global level of the text. It asked if they completely understood the text after just one reading, and how well they felt they could re-tell the content to someone else. It also invited any other comments they wished to make about the text. They were finally asked to consider what factors guided them in their overall ranking process. The reflections and opinions gathered from the questionnaire covered a range of factors, that are briefly reported on here. A copy of the questionnaire is in Appendix 9.

9.2 Grammatical structures

Some of the teachers found the grammar in the texts challenging. Pronouns and possessive pronouns were mentioned as adding to the task of gaining meaning. Low frequency fronted comments of time were also mentioned as causing difficulty, for example *nō nākuanei tonu* (just now), *auina rawa ake* (at that (I) instantly).

9.3 Polysemy

Sometimes words with the same word form have different meanings, and words which are similar to other words were mentioned as causing confusion. For example *pōnga* (night-fall) was read as *ponga* (tree fern) until the macron was attended to during a re-run, and *maunga* (capture) was read as *maunga* (mountain). This occurs when reading is still being processed heavily at the word level rather than flowing into a sentence level process to gather meaning. The same word form occurring close together with different meanings was mentioned as causing confusion, for example, *Rata* was used as a proper name and then used again as *rata* (like, prefer). Knowledge of syntax and attention to capitalisation is obviously important for sorting these things out. The use of common nouns as proper names was mentioned several times as causing a pause or re-run to check meaning.

9.4 Unknown words

Unknown words were noted as having an effect at the sentence level, but were generally able to be overcome at the global level of understanding. It was noted that a high presence of unknown words made reading tedious, slow and not enjoyable. Guessing and looking for synonyms that would fit were ways of solving unknown words. Some said they used their knowledge of other variants to test their guess. Looking for base words within words was another strategy used along with using their knowledge of grammar. The repetition of unknown words was mentioned as being helpful, and that encountering the word again in context, reinforced whether their initial interpretation was accurate or not. The surrounding words were heavily relied on for gaining meaning.

9.5 Level of understanding and ability to re-tell

While some teachers said they understood a text fairly well, this did not always translate into confidence to re-tell it to someone else accurately. Mostly, teachers felt they needed a second reading to feel confident with their level of comprehension, and ability to re-tell. This has implications for the practice of using re-telling as an assessment tool for comprehension, and would confirm that it is good practice to allow at least two exposures to a text before asking for a re-telling.

9.6 Use of idiom

The use of idiom in the texts was sometimes recorded as having caused a stopper effect because the reader had to re-run the idiom to understand it. This was especially true for the second language speakers, for example 'E kīa ana ā taihoa' (it is claimed that before long..) is not commonly used and caused problems for some.

9.7 Impact of dialect and borrowing

There were some texts that contained high use of dialectal variants. For example, *Te Kēhua o Waimā* (Hōhepa, 1974) was included because it contained a number of borrowed words typical of the speech used in the Northland area. As expected, a result of this was that one of the adult readers from the author's tribal area had an advantage in decoding this text. This was mentioned in their questionnaire feedback as being a factor which positively affected readability for them. Several younger readers or those without prior experience of this dialect, mentioned that the borrowings negatively affected readability. Borrowed and dated words which were cited throughout the whole twelve texts as having a stopper effect included: *whiro* (willow), *kiki* (gig), *tupeka* (tobacco), *wāke* (walk), *tāti* (start), *hereputu*

(hereford), *tāriana* (stallion), *pūru mātihi* (bull mastiff), *ripi tuna* (gaffing eels), *māti* (matchstick), *hōro* (shawl), *pea* (pair) and *rekoata* (record). First language speakers of Māori generally found dialect to be the only confounding feature they had to solve in reading these texts whereas second language learners mentioned these along with grammatical features and relational elements.

9.8 Reading strategies

Participants mentioned a heavy reliance on the strategy of reading back and forth to solve unknown words and parts of text to gain meaning. Another strategy was trying to get the big idea of the story by reading on. Some mentioned using knowledge of the wider context, and looking closely at the relationships of the immediate words to the unknown word/s. The strength of the introductory paragraph in setting the scene was mentioned as important for getting started with understanding and setting a direction for prediction. One teacher mentioned the strategy of thinking wider and reading beyond the lines when metaphorical language was apparent. This strategy requires a deep cultural knowledge to be applied to gain meaning.

9.9 Other comments

Most readers stated that they equated longer texts with greater difficulty, however there was also mention that this assumption wasn't always true. Some readers mentioned sentence length as being a factor they considered, and also how many ideas were contained in one sentence. Due to the natural human perception that a bigger task will be harder, total text length cannot seriously be considered as a valid criterion for estimating difficulty or text complexity. It is however a criterion which is useful in a levelling process. It is also a useful numerical figure to use in evening out the results of criteria applied to texts of varying length, and in calculating speed and error rates for reading behaviours and levels of reading fluency.

The type of vocabulary was mentioned by only two respondents as being a barrier to understanding a text. However, some readers felt that sometimes the Māori sentence structure seemed to be following English structures and this interfered with the flow of the Māori. Missing macrons was mentioned by some as a barrier. The cover sheet on the texts stated that macron discrepancies were left unaltered from the original texts unless they interfered with meaning.

Texts which readers enjoyed, they ranked as easier. Humour was referred to positively. Some readers mentioned the ease of flow of understanding without having to analyse words as being their key indicator of reading ease. This aligns with the 'stopper' theory as being a major interrupter in processing a text for meaning. Everyday dialogue was mentioned as something that made a text flow better and made it more authentic and enjoyable. Texts that readers found boring were cited as being harder to understand. These were generally the longer texts. Texts with many characters to hold in reference were also cited as challenging and boring.

Comments were made about texts which are either translations of text from English-medium material or are retellings of myths from other cultures. Stories seemingly built on European fairy tale traditions drew a mixed response. It was noted that the text *Te Ana o Te Rau* follows a similar line to the Princess and the Frog type of tale. While this type of material enriches the diversity of input that students are exposed to, there were a number of texts that were not included in the test series for those reasons.

9.10 Summary

The comments collected from the teacher questionnaire, while not used in the study for estimating difficulty, provide useful insights into the processing strategies that readers used to overcome the difficulties that a text may present. It is not within the scope of the study to discuss these findings in any depth; the purpose of the questionnaire was, as previously mentioned, a tool for strengthening teacher opinion in the ranking task.

Chapter Ten: Conclusions

The initial aim of this study was to find a simple, manageable and valid way of estimating difficulty in Māori texts written for students at the middle to senior school level. The study began by testing whether the Elley noun count method of estimating text difficulty in English texts had validity when applied to Māori texts. During the progress of this project, it became evident that there would need to be several phases to the study in order to meet the aim. Phase One of the study was developed to test the premise that nouns would be the key stopper words for young readers of Māori, as both Elley (1969) and Clay (1966) believed to be the case for English. A small study modelled on the work of Clay was undertaken in a Māori-medium setting, the findings of which, were in contrast to Clay's findings. The results did not indicate that nouns were the most problematic word class for readers of Māori. In some texts, verbs presented as more problematic than nouns. While this part of the study was too small to draw any conclusions about a particular word class dominating the stopper effect on reading, the difficulty theory about nouns in English, was not clearly supported for Māori. This was critical information for the next phase of the study which was to develop word frequency lists as Elley had done, in order to assign levels of difficulty to particular words to act as a proxy for semantic load.

In the light of the findings of Phase One, word lists were constructed to include all content word classes in the selection for the lists, not just nouns. Constructing these lists was the second phase of the project which was a major undertaking, involving the amalgamation of frequency data from large Māori language corpora. Nine word lists graded by frequency data were developed, eventually comprising 1820 of the highest frequency content words from the corpora. In order to maintain the aim of manageability, it became necessary to find computer software that could quickly analyse the vocabulary in a text in relation to the word lists. A software programme was provided by Paul Nation called *RANGE*

(Heatley et al. 2002) which sorts words in a text into lists. Setting *RANGE* to use the lists that had been constructed for this study, assisted the researcher to calculate a vocabulary load index for texts written in Māori. During the process of calculating a vocabulary index, Elley's method was further modified as it was decided to treat proper names differently to Elley. Elley's method does not include proper names in the calculation of vocabulary burden. These words are therefore assumed to be "weightless" in a text. However, studies done on the difficulty that proper names present to a reader do not conclusively support this assumption, (Hancioglu and Eldrige, 2007; Ghadirian, 2002; Kobeleva, 2005). Therefore, it was decided for this study to treat proper names as unknown words and score them as such. This was called the vocabulary index method (M1).

Formulae for estimating text difficulty in other languages generally use a combination of one semantic measure and one syntactic measure. It was therefore, decided to extend the modified Elley method and add a syntactic measure to the vocabulary index. A simple way of measuring syntactic load was proposed by Benton et al. (1995): to total the number of different function words used in a text. For this study, this was done by developing a list of function words for *RANGE* to use in its analysis which provided a quick way to total the number of different function words in a text. The vocabulary index for a text was then multiplied by the number of function types in the text to produce this combined method.

These two phases complete, the study had now progressed to the point that the linguistic analysis tools were in place to rank texts using two methods, ie a semantic measure of vocabulary burden alone and a second method using a combined semantic and syntactic measure. Both of these methods evolved from the early work of Elley (1969) and Benton et al. (1995). A comparison was made of text rankings using the combined method (M2) with rankings made by Benton et al. (1995) in the *Ngā Kete Kōrero Framework Project*. This comparison

produced correlations of 0.93, which showed that with a syntactical measure built in to the method the findings of this study were consistent with previous work in this field.

The third phase involved applying these methods to select and rank a series of texts to take to Māori-medium settings to validate. From an analysis of 33 texts, two series comprising six texts each were selected. One series was selected using M1; the other was selected using M2. This created two different series for validating the methods of calculating text difficulty with the criterion measures.

The study now required criterion measures with which to validate these methods. Criterion measures were established in phase four of the study by gathering opinions from teachers and students involved in the field of Māori-medium education. They were asked to rank the six texts in order of difficulty. Students' reading comprehension achievement on test questions based on each text provided additional measures. As supported by other studies, (Elley & Croft, 1989; Harrison, 1980; Klare, 1974; Maxwell & Benton, 1995) a criterion measure of combined teacher and student opinion was found to be a reliable way of validating readability formulae. Although it would have been preferable to have included a large number of texts in the validation, ranking just six texts proved to be a challenging task for participants.

Further to the ranking task, a teacher questionnaire was included in order to focus teacher reflection on particular features of the texts and to assist them to make their ranking process tighter around those aspects. In spite of the small number of texts (12 in total), useful information on qualitative aspects affecting text difficulty in Māori was produced. Findings from the questionnaires and comprehension questions confirmed that prior experience, coupled with high interest in content, play a part in determining ease of readability. These are aspects that will always remain immeasurable and specific to individual readers.

The role that unknown words play in creating a stopper effect on reading and comprehending was also confirmed showing that the number of unknown words in a text increases the comprehension burden in Māori as it does in other languages. Data also indicated that proper names should be included when calculating a vocabulary load score; total text length strongly affects students' presumptions about text difficulty; irregularity of macronisation; and dialectal spelling variants also contributed to text difficulty.

In phase five, the methods were validated using a Spearman's Rho rank order analysis with the criterion measures. The relative success of a method was determined by the strength of the correlation coefficient with the criterion measures and its consistency across both series of texts. It was found that the make up of the two series was very different. One of the series drew fairly consistent results when different methods were applied to it but the other seemed to be consistently at odds with itself showing dramatic changes in ranking when different methods were applied. The issue of analysing how well built a text is overall, emerged during this process. In other words, the difficulty of vocabulary should be tied to the difficulty of syntax, to the average sentence length, and to overall text richness or lexical density.

It was found that using the modified Elley method, which is a semantic measure of vocabulary burden alone, produced no correlations of statistical significance with the criterion measures, for either series of texts. However, in contrast to this, the second method (M2) which added Benton's syntactic measure to the formula, produced strong correlation coefficients with criterion measures for both series of texts. This finding is in contrast to those of Elley & Croft (1989) who determined that for English, combining function words into the analysis added considerably to the complexity of the process but showed no increase in validity. This was an important finding, which suggests that function words (which were found to make-

up on average, 62% of the texts) play a very significant role in determining text difficulty in Māori.

Once this analysis was completed, it was decided to examine the validity of a variety of other methods of estimating text difficulty. A further five ways of estimating text difficulty were explored. Since the combined semantic/syntactic method tested (M2), showed such an improvement in validity with the inclusion of the number of function word types, it was decided to test the strength of simply counting the number of different function words used in the text (M3). This method was found to have slightly stronger correlations with the criterion measures for both series than the combined method (M2). Average sentence length was also tested as a syntactic proxy on its own (M4), but only produced significant correlations with one of the series of texts. The next method to be examined was type:token ratio (M5) which was also a component of Benton et al's. (1995) formula. Type:token ratio gives a measure of the lexical richness of a text. Studies by Richards & Malvern (1997) assert that type:token ratios are not independent of sample size and so this measure was standardised to account for the varying text lengths that both series contained. The software programme *WordSmith Toolsv4* (Scott, 2004) was used to calculate the mean standardized type:token ratio. However, this method also only produced statistically significant correlations with the criterion measures for one series. The semantic measure using a vocabulary index was further tested in combination with average sentence length as the syntactic proxy (M6). This method only produced strong correlations with one of the series of texts. Finally was decided to try a combination of the number of function types with the mean standardised type:token formula (M7). This produced statistically significant results for both series.

These findings brought three methods to the forefront of the validation process; vocabulary burden with function types (M2), function types alone (M3), and, mean standardized type:token ratio with function types (M7). Assuming that it is preferable to measure a text by the simplest means, yet also in the most global way, the two formulae which use both semantic and syntactic proxies, would be preferred. They also align with basic measures used in readability formulae used for other languages. In deciding which of these two formulae should be considered more robust, it was necessary to consider carefully the information they produce about semantic load. The vocabulary burden theory represented by a vocabulary index using word frequency lists (M2), gives an indication of the kinds of words and the levels of familiarity that will be encountered in a text. However, this measure will not show how much repetition occurs in a text to assist the reader. On the other hand, the lexical richness theory represented by the type:token ratio formula will give an indication of repetition provided in the text, but will not show how difficult or unfamiliar the vocabulary is likely to be. Both formulae include the same syntactic proxy of number of function types. This leads to the conclusion that a formula combining all three: type:token ratio, function types and a rating of vocabulary need to be tested further. These three features of texts have been shown in this study to be the key indicators of text difficulty. The study highlights for teachers that having a “quick look” at a text to match it to a student’s capability can be very deceiving.

Emergent issues:

There were a number of unforeseen issues which emerged during the course of the study, these usually presented as limitations which will need to be taken into account for further research. They are discussed below.

Corpora

There is a critical limitation to all of the corpora used in this study because they are all representative of the speech and writing of adults. Elley sought to find

appropriate word frequency data for his word lists by eliminating those which contained primarily adult writing. Cedric Croft's (1983c) project to establish a corpus of children's writing in English was an intervention designed to alleviate this problem. The *Māori Broadcast Corpus* (Boyce 2006), *Ko Ngā Kupu Pū Noa*, (Benton et al. 1982 & 1983), and the *(Huia) Māori Children's Text Corpus* (Huia, in progress), are all collections of predominantly adult productive language.

In addition, it is accepted that the accuracy of word frequency data used for constructing graded word lists, is dependent upon the size of the corpora. Richards (1975) says that the extent to which true frequencies of occurrence of relatively uncommon words can be measured depends very strongly on the size of the corpus, which inevitably has to be very large. While the corpora used for this study were the largest available at the time, the limitations of corpus size needs to be acknowledged. Furthermore, an added constraint presented by the *(Huia) Māori Children's Text Corpus*, is that it is still in an unfinished form and as such, hasn't yet been shaped and balanced to meet certain criteria. For example, there are some very long texts included in the corpus which over-represent styles of particular authors and the topics of those texts. There are also references to publishers, authors and series which are still included in the corpus, and are therefore, represented in the frequency data. This latter issue was intercepted by examining the data manually where some words were showing unusually high frequencies, however, this was not always noticed and was very time consuming. In addition, a number of early level texts are not yet included in the corpus which lowers the frequency and range data of some words known to be high frequency words encountered in early reading material. Difficulties also arose because the words included in the *CMTC* texts have not been through a spelling standardization process. This meant that many words were represented more than once in the data in different forms, and their data had to be put back together manually. Notwithstanding these limitations, because the application of the readability formulae tested in this study were designed to grade the very

same material that makes up the *CMTC*, it was still considered the best available corpus to use for this study.

Word lists

The best way of constructing word lists is a topic of ongoing debate. The word lists made for this study were organic in their development and are constantly under review as questions about the placement and groupings of words continue to arise during text analysis. However, at some point a decision needs to be made about word inclusion, frequency and range data cut off points, and whether or not to allow expert opinion to influence objective data. Decisions were made concerning all of these aspects during the construction of the lists and different decisions would have produced different lists and possibly different results. In order to strengthen the definitions of frequency values for word frequency lists, the addition of a children's productive language corpus to further complement the current pool of corpus material is essential.

The nine word lists made for this study were guided by the amalgamation of all currently available corpus material relevant to an educational context, and they were produced with careful alignment of a range of data supporting the placement of words. Notwithstanding this process, the decisions about the cut off points used to divide and grade the whole collection into sections and also the size of the corpora used are limitations that the lists may have in their present form.

A key finding from this part of the project is that constructing word lists is in itself a complex process, and requires time for the development of deeper thinking about the best way to truly represent the frequency of use of words and therefore word knowledge amongst second language learners of Māori. Polysemy presented a difficulty in constructing the word lists and in the analysis process.

The high number of spelling discrepancies that exist in the texts as they have developed over a forty year period was also found to be problematic for analysis purposes and remains so for teachers. An attempt was made in this phase to apply not only computer generated data in the process of making the lists, but also to acknowledge and incorporate the intuitive knowledge that teachers of the Māori language have built up over many years. To this end the lists represent a quantitative and qualitative approach in their construction.

When Benton et al. (1995) undertook their early analyses of texts using word lists, they used a software programme known as *Vocab profile*. This was the precursor software to *RANGE* which was used in this study. It seems that the *Vocab Profile* software for text analysis in Māori has lain dormant since 1995. Reviving the application of this software for Māori, and making word lists derived from frequency data for *RANGE* to use, has in itself been a worthwhile outcome of this project. The advance in using *RANGE* to analyse the texts has wide ranging implications for the ongoing development of texts written in Māori for children, and for the analysis of children's writing. Elley (1969) previously described the analysis of whole texts as laborious. However, using *RANGE* brings whole text analysis within the reach of any practitioner. The information that can be gained about the vocabulary and structure in a text will be worthwhile as a teaching tool, in addition to providing rich and detailed information for a levelling process.

This phase of the study revealed coverage data for Māori texts which has previously not been available. It was shown from the snapshot of the 36 texts analysed for this study, that coverage provided from just over the first 1000 words in Māori averaged 92%, and ranged from 81% to 96%. It can be seen that this type of analysis is now possible to assist in levelling processes. Hu and Nation (2000) recommend that 98% of the words in a text should be 'known words'. Recommended rates of known words that a text should contain for readers of

Māori are as yet not determined, and the coverage statistics of the available reading material has not been produced.

A further unresolved issue in the construction of word lists is that of grouping words into families. Investigation into the role that derivatives and affixed forms of words play in familiarity and true word knowledge is required, such as that which was described by Bauer and Nation (1993) for English.

In spite of the difficulties that arose, the researcher is confident that the decisions made were appropriate for the challenges that arose at the time and for the scope of this project. Only further investigation into this area will inform us otherwise.

The conflicting nature of the text series

The conflicting results produced by having selected two separate series of texts was unforeseen, but not necessarily unfortunate. By having a “rogue” series of texts, the challenge was put forward to find a formula that would produce statistically significant results regardless of the disparities in text make-up. While this was initially considered to be a flaw in the methodology, it was eventually shown to be a strength. The wide range of factors that can provide challenge in the make-up of a text were brought to the fore. It was shown that combinations of vocabulary load, sentence length, lexical richness, and syntactic complexity, need to be tied together in a close range of difficulty to produce a well balanced text. In the future, these components can be analysed separately into a profile for a text to provide authors with a sense of achieving a good balance in a text. The findings during the final phase were complicated by the instability of one series of texts and as mentioned earlier, it was tempting to disregard the results that this series kept producing in favour of the stability shown by the other series. However, since there are always going to be texts which are not as well balanced as others, it was best to find some middle ground and look for the methods that

showed the strongest validation for both series. Eventually, only the methods that could produce statistically significant results for both series, were considered most valid.

Participant numbers and range of texts

As with many projects of this size, at the time of conclusion, the researcher is left wishing that a bigger number of participants were included and that a larger sample of texts was tested. Both of these aspects have been constrained by the scope of this study. However, in future studies of this type, a way of gathering a larger student voice, and more teacher opinion over a wide range of texts would be sought in the initial design of the methodology. One solution to this could be to use texts and student performance results from data which is already collected on a national basis such as that from Assessment Tools for Teaching and Learning (AsTTle). A larger study with a national approach would require sponsorship by interested organisations.

Tagging software

At this time, there is no tagging software available which can organise the words of a Māori text into word classes, as there is for texts of other languages. The main value of having tagging software available for this study, would have been to trial a true noun count method. Analysing texts manually into word classes proved to be a laborious task and therefore at this point the noun count method would not be a useful or manageable formula to apply. Private communication with the writer of the software for the noun frequency count method for English texts, revealed that a barrier to classroom teachers using this method, was that the nouns must be manually selected from a passage. While at the production level for English texts this is made manageable by the use of tagging software, this is not available to Māori-medium at this time.

Should tagging software be developed for the Māori language, new opportunities will arise for gaining insights into the use of the language by speakers of all ages and backgrounds and about how the Māori language is developing and changing over time.

Future directions

This study has highlighted the need for further research that would support the continued development of readability measures and other aspects of research development in literacy for the Māori-medium sector. In addition, it would be important that Māori researchers themselves be supported through resources and time to undertake such research and development as the background pool of knowledge about the language is still in a phase of developing. The most immediate future needs identified by this study are outlined below.

Corpus development

The need for corpus development for Māori is at a critical point in New Zealand at this time. The Ministry of Education is poised to implement National Standards in all English-medium and Māori-medium primary schools, and is embarking on describing spoken language progressions that should be achieved by students in Māori-medium settings at specific times in their schooling. It is recognised that oral language is the platform from which all aspects of literacy develop, yet there is still no body of spoken productive language established that can represent this fundamental platform accurately for the Māori language. Without a corpus of children's productive language, it will be very difficult to describe natural language pathways taken by young learners of the Māori language. Furthermore, it will be difficult to produce such descriptors with any confidence in their appropriateness or accuracy. In addition, later learners of the language are likely to follow slightly different pathways of language development. A well designed corpus of productive language would be able to describe many things about the state of the language amongst the current generation of speakers. For example,

patterns of language from different cohorts in immersion settings, language and vocabulary progressions, regional patterns and dialectal usage, common grammatical difficulties, needs for vocabulary growth, and word frequency. This is the key direction for developing new knowledge about the Māori language at this time.

Once a corpus of children's productive language in Māori (spoken and written) is developed, it will become a fundamental body of evidence from which many other developments, as determined by the Māori-medium sector, will emerge.

Further testing

It is obvious that the three key aspects of measuring text difficulty found to be valid in this study are in need of further testing on the text material that is available to middle and senior school students in Māori-medium education. It is desirable that a formula that produces a linguistic profile of a text is applied to a large body of texts in order to rate them according to processing burden. Furthermore, in reference to the understanding that these measures must not stand alone, wider criteria would then need to be applied to modify the final difficulty ratings given to such texts. It is now important to carry out further research using a larger number of texts, and to involve a larger number of student participants. New ways of rating need to be explored in order to establish criterion measures across more texts or perhaps reader performance could be used to produce a rank order as used in early research on readability by Bormouth (1966) and Elley (1982). The results in this study showed student performance to be a reasonably good measure of text difficulty that could realistically be pursued further. Producing ranking data for a larger number of texts will enable validation analyses for several methods to be repeated.

In Conclusion

The findings of this project show promising trends towards being able to estimate the processing burden a Māori text carries. Proxies for three key features of text have been found to be valid for determining text difficulty for Māori; type:token ratio, function word types, and a measure of word familiarity. There is however, still a need for further validation across a larger field of texts, students and teachers. The project has shown some validation of Benton et al's.(1995) early formula which was applied to texts but not tested in the field. Useful direction has been signalled for further research, and information which was previously unknown about Māori texts written for children has been revealed. The word lists have many possible applications outside of this project, and it is hoped to further their development and use. Above all, a good start has been made that will enable an analysis of Māori texts by resource developers and publishers at the production level.

References:

- Allyn and Bacon. (1977). Semantic feature analysis; An interactive strategy for vocabulary development and text comprehension, Special Issue on Vocabulary. *Journal of Reading* 29 (7) International Reading Association. United States of America.
- Anderson, J. (1971). *A technique for measuring reading comprehension and readability*. 25, 2 English Language Teaching: London.
- Anderson, Richard C., and Peter Freebody. (1981). *Vocabulary Knowledge*. In Comprehension and Teaching; Research Reviews, John T. Guthrie. (Ed). Newark, Del; International Reading Association.
- Bauer, W. (with W. Parker, T. Evans & T. Teepa.) (2003). *The Reed reference of grammar of Māori*. Auckland: Reed.
- Bauer, L., Nation, P. (1993). Word Families. *International Journal of Lexicography*, Vol 6, No .4. Oxford University Press.
- Benton, R., Glynn, T. , Berryman, M. , Kapa, A. , Murphy, H. , Hindle, R. , Rau, C. (1995). *Ngā kete kōrero policy project*, A report to Te Puni Kōkiri. Wellington: Te Puni Kōkiri.
- Benton, R.A., Tumoana, H. and Robb, A. (1982). *Ko ngā kupu pū noa o te reo māori*. Wellington: New Zealand Council for Educational Research.
- Benton, Richard Anthony. (1982) (a) The anatomy of a word list. *SET: research information for teachers, no.2*. Wellington: New Zealand Council for Educational Research.

- Benton, Richard Anthony. (1983). *Basic Māori word list: levels six to ten*.
Wellington: New Zealand Council for Educational Research.
- Berryman, M., Rau, C., Glynn, T. (2001). Ngā kete kōrero: A framework for
assigning levels of difficulty to existing and new Māori reading resources.
SET Research Information for Teachers 2001;2 Wellington: New Zealand
Council for Educational Research.
- Bishop, R., Berryman, M., Richardson, C. (2001). *Te toi huarewa. Effective
teaching strategies and effective teaching material for improving the
reading and writing in te reo Māori of students aged five to nine in Māori-
medium education*. Draft report to The Ministry of Education.
- Biggs, B. (1969). *Let's learn Māori: a guide to the study of the Māori language*.
Wellington, Auckland: Reeds.
- Bormuth, J.R. (1966). 'Readability: A new approach', *Reading Research
Quarterly*, 1, (3), 79-132.
- Boyce, M.T. (2006). *A Corpus of Modern Spoken Māori*. Phd Thesis. Victoria
University of Wellington, Wellington, New Zealand.
- Clay, M. (1966). *Emergent reading behaviour*. Unpublished doctoral
dissertation. University of Auckland
- Clay, M. (1993). *The early detection of reading difficulties 3rd Edition*. Auckland:
New Zealand. Heinemann.

- Clay, M. (2007). *An Observation Survey of Early Literacy Achievement 2nd Edition*. Auckland: New Zealand. Heinemann.
- Crain, S., and Shankweiler, D. (1988). Syntactic complexity and reading acquisition. In A. Davidson and G.M. Green (Eds.), *Linguistic complexity and text comprehension; Readability issues reconsidered*. Hillsdale, NJ: Lawrence Erlbaum.
- Davis, Frederick B. (1944). Fundamental factors of comprehension in reading. *Psychometrika* 9; 185-97.
- Dale, E. and Chall, Jeanne S. A (1948). 'A formula for predicting readability'. *Educational Research Bulletin*, 27, 11-20.
- Davison, A. (1985). *Readability – the situation today*. Technical Report No 359, Champaign; Centre for the Study of Reading.
- Education Review Office. October (2000). *Literacy Education in Kura Kaupapa Māori*. Wellington: Ministry of Education.
- Elley, W., Croft, C., Cowie, C. (1977). *A New Zealand basic word list, revision of the Dolch basic sight vocabulary*. Wellington: New Zealand Council for Educational Research.
- Elley, W.B. (1982). 'The Difficulty with Prose'. Unpublished paper presented at Conference of NZARE in Elley, W.B. and Croft, A.C. (1989). *Assessing the difficulty of reading materials: The Noun Frequency Method Revised Edition*. NCER: Wellington.

Elley, W.B. & Croft, A.C. (1989). *Assessing the difficulty of reading materials: The Noun Frequency Method Revised Edition*. NCER: Wellington.

Elley, W.B. & Smith, J.W.A. (1994). *Learning to read in New Zealand*. Auckland: Longman Paul.

Elley, W.B. & Smith, J. (1997) *How children learn to read. Insights from the New Zealand experience*. Auckland: Longman.

Ellis, R. (2005). *Instructed second language acquisition, A literature review*. Report to the Ministry of Education. Auckland: Auckland Uniservices Limited.

Flesch, R. Flesch-Kincaid readability test. Retrieved February 7, 2008 from: http://en.wikipedia.org/wiki/Flesch-Kincaid_Readability_Test

Fry, E. Fry Readability Formula. Retrieved February 7, 2008 from: http://en.wikipedia.org/wiki/Fry_Readability_Formula

Fry, E. (2002). Readability versus leveling. *The Reading Teacher*; Nov 2002; 56, 3; Academic Research Library.

Froese, V, & Kurushima, S. (1979). The effects of sentence expansion practice on the reading comprehension and writing ability of third graders. In M.L. Kamil & A.J. Moe (Eds), *Reading research: Studies and applications*, twenty-eighth yearbook of the National reading Conference. Rochester, NY: National Reading Conference.

- Garcia, G. (2003). The reading comprehension development and instruction of English-language learners. In A P Sweet & C.E. Snow (eds.), *Rethinking reading comprehension*. (pp33-50). New York: Guilford Press.
- Glynn, T., Harawira, W., & Durning, C. (1993). *Tatari, Tautoko, Tauawhi: Hei Āwhina Tamariki ki te Pānui Pukapuka*. Tauranga: Special Education Service.
- Glynn, T., Berryman, M., Bidois, P., Furlong, M., Thatcher, J., Walker, R., & Atvars, K. (1996). Bilingual gains for tutors and tutees in a Māori reading programme. *He Paepae Kōrero. Research Perspectives in Māori Education*. Wellington: New Zealand Council for Educational Research.
- Hancioglu, N. and Eldridge, J. (2007). Texts and frequency lists; some implications for practicing teachers. *ELT Journal Volume 61/4 October*. Oxford University Press.
- Harlow, R. (2001). *A Māori reference grammar*. Auckland: Pearson Education, Longman.
- Harrison, C. (1980). *Readability in the Classroom*. Cambridge:Cambridge University Press.
- Heatley, A., Nation, I.S.P. and Coxhead, A. (2002). *Range and Frequency programmes*. http://www.vuw.ac.nz/lals/staff/Paul_Nation
- Heimlich, J.E. & Pittelman, S.D. (1986). Semantic mapping: Classroom applications. *International Reading Association*. Delaware: USA.

Holdaway, D. (1979). *Foundations of literacy*. Ashton Scholastic: Gosford N.S.W.

Hu, M. & Nation I.S.P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*. 13.

Hudson, R. (2007). About 37% of word-tokens are nouns. *Language* 70, pp 331-339, 1994 Retrieved on 14th April 2009 from <http://www.phon.ucl.ac.uk/home/dick/37%25.htm>

Huia Publishers (2005). *Te Kete Kupu*. Huia Publishers: Wellington.

Ingram, D.E. and Elias, G.C. (1974). Bilingual Education and Reading. *RELC Journal* Vol. V no. 1 June.

Klare, G.R. (1963). *The measurement of readability*. Ames, IA; Iowa State University Press.

Lorge, I. (1944). 'Predicting Readability'. *Teachers' College Record*, 45, 404-419.

May, S. Hill, R. Tiakiwai, S. (2004). *Bilingual/Immersion education: Indicators of good Practice*, final report to the Ministry of Education. Wilf Malcolm Institute of Educational Research: University of Waikato.

Maxwell, I. T. and Benton, R.A. (1995). *Te Kura Ki Uta: Opotiki College Bilingual Education Programme Assessment and Evaluation Project. Final Report and Resources*. Te Puni Kōkiri: Wellington:

1. Word frequencies in classroom reading materials.
2. He aha ngā kupu? Text analysis programme and guidelines.
3. Word index to Māori reading materials.
4. Guidelines for monitoring and facilitating vocabulary development.

Metametrics, The Lexile Framework, Retrieved on February 7, 2008 from:
<http://www.Lexile.com>.

McCarthy, D. (1954). *Language development of children*. Cited in Barham, I. (1965). *The English vocabulary and sentence structure of Māori children*. NZCER: Wellington.

Ministry of Education. (1999). *Aromatawai Urunga-ā-kura: AKA/SEA*
Wellington: Learning Media.

Ministry of Education. (2000). *Using running records, A resource for New Zealand classroom teachers*. Wellington: Learning Media.

Ministry of Education, (2008). *English Language Learning Progressions*.
Wellington: Learning Media.

Moyle, D. (1970). *Readability - the use of cloze procedure. Reading and the Curriculum*. (ed. John Merritt.) Proceedings of 7th Annual Study Conference of U.K. Reading Association: Ward Lock International.

Nation I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation I. S. P. Ed. (1996). *Vocabulary lists*. Victoria University of Wellington.
English Language Institute Occasional Publication No. 17

Nation, P. and Worthington, D. (1996) *Using texts to sequence the introduction of new vocabulary in an EAP course*. RELC Journal Volume 27, 2: pp 1-11.

- New Zealand Ministry of Education. (2007). *Te Reo Matatini: Māori-medium literacy strategy*. Wellington: Huia Education.
- Nicholson, T. (1991). *Overcoming the Matthew Effect: Solving reading problems across the curriculum*. Wellington: New Zealand Council for Educational Research.
- Nicholson, T. (1997). *Solving reading problems across the curriculum*. Wellington: New Zealand Council for Educational Research.
- Rau, C. (1998). *He matai āta titiro ki te tutukitanga matatupu, pānui, tuhi: The Māori reconstruction of an observation survey of early literacy development by Marie Clay*. Ngāruawahia: Kia Ata Mai Educational Trust.
- Rau, C. (2001). *A Framework for Capturing the Literacy Achievement of Year One Students in Māori-medium*. Paper presented at Assessment Hui for the Ministry of Education, Wellington, September 2001. Retrieved 28 January, 2008. from: <http://www.kiaatamai.org.nz/angitu/index.htm>
- Rau, C. (2004). *He Mātai Mātātupu: Assessment for Māori-medium Literacy Learning*. Waikato University.
- Rau, C. and Mc Naughton, S. (1998). *Reconstructing the tasks in Māori*. In C, Rau (Ed) *He matai ata titiro ki te tutukitanga matatupu, pānui, tuhi. Pukapuka matua. The Māori reconstruction of an observation survey of early literacy achievement by Marie Clay*. Ngāruawahia: Kia Ata Mai Educational Trust.

- Riley, P. M. (1973). *The cloze procedure – A selected annotated bibliography*. Lae: Papua New Guinea. University of Technology.
- Richards, J.C. (1974). Word lists: problems and prospects. *RELC Journal*, Vol. V, no. 2 . SEAMEO Regional English Language Centre: Singapore.
- Richards, B., & Malvern, D. (1997). Quantifying Lexical Diversity in the Study of Language Development. *The New Bulmershe Papers*. University of Reading: Reading.
- Rye, J. (1982). *Cloze procedure and the teaching of reading*. London: Heinemann Educational Books.
- Saville-Troike, M. (1984). What really matters in language learning for academic achievement. 18.2 *TESOL Quarterly*.
- Simon, J. and Smith, L .T. (2001) *A Civilising Mission? Perceptions and Representations of the New Zealand Native Schools System*. Auckland University Press: Auckland.
- Scott, M. (2004). *Oxford WordSmith tools version 4.0* . Oxford: Oxford University Press (Computer Programme, manual and associated files).
- Stahl, S. (1986). Three principles of effective vocabulary instruction, *Journal of Reading, Special Issue on Vocabulary*. 29 (7) International Reading Association. United States of America.

- Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, 56, 72–110. In Bialystok, E. (2007). *Acquisition of Literacy in Bilingual Children: A Framework for Research*. *Language Learning* 57: Suppl. 1, June 2007, pp. 45–77.
- Te Pou Taki Kōrero. (1996). *Ngā kōrero tawhito o Te Wharekura*. Wellington: Te Pou Taki Kōrero.
- Thomas, E. & Robinson, H. (1977). *Improving reading in every class 2nd Edition*. Boston.
- Thomson Nelson, (2003). *Key to the PMs: The Levelling of PM books and Benchmark kits*. Thomson Learning Australia, Nelson Price Milburn. Petone: Wellington.
- Thompson, G. B. & Nicholson, T. Editors. (1999) *Learning to read beyond phonics and whole language*. New York: Teachers College Press.
- Underwood, B.T., & Schulz, R.W. (1960) *Meaningfulness and verbal learning*. Lippincott. Chicago.
- White, S., Clement, J. (2001). *Assessing the Lexile Framework: Results of a Panel Meeting*, NCES 2001-08, U.S. Department of Education, National Center for Education Statistics. Washington, DC.
- Vincent, D., Pugh, A.K., Brooks, G. (Editors). (1986). *Assessing reading proceedings of the UKRA colloquium on the testing and assessment of reading*. London: Macmillan Education Ltd.

Wallace, D. (2003) The Development of an NCEA Corpus and Word List

TESOLANZ. Retrieved on February 10th 2008 from:

http://72.14.253.104/search?q=cache:qmVFUuKwTjMJ:www.english.unitechnology.ac.nz/nesb/resource_exchange/resources/NCEAWordListIntroduction.doc+word+lists&hl=en&ct=clnk&cd=5&gl=nz

Zakaluk, B.L., & Samuels, S.J. (Eds.) (1995). *The Lexile framework for reading*. Durham, NC: Metametrics.

Appendices

List of Appendices

1. Phase One texts and comprehension probes used in the Clay replica study for the <i>Kete Pīngao i</i> level, <i>Kete Pīngao o</i> level and <i>Kete Miro</i> level	1
2. Phase Two, words lists organised in score sets 1-9.....	10
3. Phase Two, function word list	56
4. Phase Two, Variant spellings and dialects occurring in texts and subsequently incorporated into <i>RANGE</i> Baseword Lists 1-10.....	60
5. Phase Three, table of 33 texts with data.....	65
6. Reference list of all Children's texts used in the study.....	67
7. Phase Four, <i>RANGE</i> data for both series of texts.....	69
8. Phase Five, Student performance tasks: multichoice and vocabulary check for both series of texts.....	158
9. Teacher questionnaire and teacher and student ranking sheets.....	182