Libraries and Learning Services

University of Auckland Research Repository, ResearchSpace

Copyright Statement

General copyright and disclaimer

First Language Reflection
on Second Language Writing

Impact and Implications

Anne S Moir Scott

ABSTRACT

In New Zealand's high-stakes school assessment system, the National Certificate of Educational Achievement (NCEA), foreign language writing skills are measured in internally assessed portfolios which comprise texts in the target language set and marked by the teacher. Because the writing is typically collected over time, and feedback can be given, portfolio assessment offers students the opportunity to reflect on their progress. However, in a crowded curriculum where time is tight, teachers tend to omit unassessed components, such as reflection, from their programmes. This thesis details a mixed method, counter-balanced, comparative study that investigated whether the omission or inclusion of a dedicated reflection component in assessment portfolios makes a difference to students' writing performance and language learning. The study was conducted in five secondary school French as Foreign Language classrooms. It addressed the question of the potential impact of reflection by measuring and comparing the learning gains in the second language (L2) writing of the 71 Year 11 participants when they reflected and when they did not. It also sought insights from the first language (L1) reflections in terms of language learner strategies and self-regulation. For the purposes of analysis, the participants, who were already placed in intact classes by their schools, were also placed for closer research analysis into three pre-test groups, Achieved, Merit or Excellence[1], using an adapted NCEA rubric.

Using both quantitative and qualitative tools to analyse the L2 data collected, the research found that the Task 1 Merit intervention participants appeared to benefit from the reflection intervention with a statistically significant gain in L2 written proficiency and L2 accuracy, but not in complexity, fluency, verb error elimination or verb self-regulation. The L1 reflections were also analysed using quantitative and qualitative tools. The research found a strong correlation between the participants' L2 achievement levels and the effectiveness of the strategic and self-regulatory practices that they reported using. Previous scholarship has found that different language features develop at different rates for different students, and that, when strategies are taught, achievement levels rise. It is possible, therefore, that providing the students with the time

---

[1] NCEA is a standards-based assessment. When students present work to meet each standard, they are awarded one of four grades for that standard, Not Achieved, Achieved, Achieved with Merit or Achieved with Excellence, depending on how well they meet the assessment requirements.

and support they needed to reflect in their L1 as they wrote for L2 portfolio assessment focussed their attention in a such a way that language learning increased for those students who were at a point in their Zone of Proximal Development (ZPD) where they were ready to take next steps in interlanguage development related to L2 written proficiency and accuracy. It seems that they needed more explicit support, in addition to reflection, before being able to develop other language features. For other participants, however, there were no statistically significant shifts. It is possible that, if the research had continued for a longer period and included regular opportunities for developing a habit of reflection, all students might have benefitted with increased or accelerated learning gains and that this learning might have included other language features in addition to L2 written proficiency and accuracy. It is important to note that the L1 reflections provided data on individual learning that had the potential to inform teaching practice. They also provided an opportunity for students to converse with the teacher in such a way that productive learning relationships were able to develop. As a consequence of the analysis of the L1 and L2 data, this study draws attention to the strong links between Assessment *for* Learning, writing to learn language and reflective self-regulatory practices.

DEDICATION

To my valued colleagues,
Aaron, Alistair, Estelle, and Helen,

four hardworking, conscientious French as Foreign Language teachers,

whose unconditional positive regard for their students and whose lifelong love of languages

continue to inspire and motivate peer and protégé alike.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

TABLES

FIGURES

GLOSSARY

BYOD	Bring Your Own Device
CAF	Complexity, Accuracy and Fluency
CEFR	Common European Framework of Reference
CLT	Communicative Language Teaching
EFL	English Foreign Language
ERO	Education Review Office
ESL	English Second Language
FFL	French as Foreign Language
L1	First Language
L2	Second Language
L3	Third Language
MOE	Ministry of Education
NZAFT	New Zealand Association of French Teachers
NCEA	National Certificate of Educational Achievement
NZCER	New Zealand Council for Educational Research
NZC	New Zealand Curriculum
NZQA	New Zealand Qualifications Authority
SLA	Second Language Acquisition
SOLO	Structure of Observed Learning Outcomes
TKI	Te Kete Ipurangi (The Knowledge Basket)
US/USA	United States/United States of America
ZPD	Zone of Proximal Development
VCE	Victorian Certificate of Education

CHAPTER ONE: INTRODUCING MY RESEARCH

Writing is a valuable tool for assessing second language learning; it is also a valuable tool for learning a second language (Manchón, 2011a; Reichelt, 2011). Second language (L2) writing researchers have been particularly interested in how writing promotes language learning and many have offered insights for creating a blueprint of the L2 writer (Schoonen, et al., 2009). I have long held the belief that reflective practices play an important role. I also believe that, when the student's level of L2 knowledge is low and the cognitive load of the task is heavy, dedicated time needs to be set aside for reflection, and the reflection will most likely occur in the first language (L1). My research therefore investigated whether L1 written reflection increases L2 learning when adolescents write for portfolio assessment.

This chapter introduces my work in four sections. The first section explains the issues that led to my research and why I believe the research to be important. The second section describes the local context, the setting for the research: the New Zealand secondary school foreign language classroom where there are diminishing numbers of students, where Communicative Language Teaching (CLT) approaches are prevalent, and where progress is assessed via portfolio. The third section establishes the global context, the theory that frames the research, and, in the process, defines the terms used throughout the document. The fourth section summarises the present chapter and introduces the seven chapters that follow.

1.1: Concerns Arising from My Teaching Practice

In December 2013, I completed my first year of doctoral studies and my fortieth year as a teacher. During those years of educating others as well as myself, I have taught first, second and foreign language English as well as foreign language French; in New Zealand, French Polynesia and the United States; to primary, secondary, and tertiary learners; and in face-to-face, online and blended classrooms. Questions related to my teaching practice have often initiated informal classroom research. Three of these issues have led to the doctoral research detailed here.

1.1.1: Three Issues

The first issue is the diminishing number of students entering senior secondary school foreign language classes in this country. Over several decades I have asked my students why they chose

to continue or discontinue learning foreign languages. I wanted to know whether there was anything more my colleagues and I could do to improve retention rates.

The second issue concerns higher order thinking, metacognition and self-regulation, what they look like in foreign language settings, and whether it is worthwhile taking time to teach them. Since the 2007 introduction of the revised New Zealand Curriculum (NZC), and consequent alignment of high-stakes assessments with the aims of the revised NZC through the National Certificate of Educational Achievement (NCEA) system, there has been an increase in assessment by portfolio in all areas of learning. In fact, portfolios have become a common component of twenty-first century secondary school programmes world-wide. According to Black & Wiliam (1998), "[a] portfolio is a collection of a student's work, usually constructed by selection from a larger corpus and often presented with a reflective piece written by the student to justify the selection" (p. 45). As well as using reflection to aid the selection of showcase pieces, students may, in the context of L2 writing portfolios, be asked to reflect on the feedback they receive in order to improve their writing across multiple drafts (I. Lee, 2007). However, I have learned from my own classroom practice that, unless students are allocated class time to do this, many do not make the effort. In addition, anecdotal evidence from colleagues at professional development events and in listserv[2] conversations suggests that, in order to maximise time, New Zealand language teachers tend to omit from their portfolio programmes unassessed components such as reflection, and focus on the writing to be assessed. Zhang (2013) found a similar situation in Hong Kong where teachers were loath to include unassessed metacognitive elements in their crowded, high-stakes programmes. It must be asked, therefore, whether setting aside precious time for students to reflect when they write actually matters.

Alongside diminishing numbers and the dropping of reflection from writing portfolios, the third issue arose from a 2012 Education Review Office (ERO) publication. The role of the ERO in New Zealand is to evaluate and report on the education delivered in schools. This body reported (ERO, 2012) that less than a third of New Zealand schools used assessment data well; it stated that effective teachers regularly collect data in order to give students specific feedback on their progress: they let their students know what they have learned and what they still need to learn,

---

[2] The New Zealand Association of French Teachers (NZAFT) listserv is a lively email exchange managed by the NZAFT Executive; French teachers around the nation, who are often the sole French teacher in their school, use it to ask questions, share resources, and give and receive information. As the NZAFT Auckland chairperson, 1999-2000, I was instrumental in setting up its first iteration.

and they help their students manage their own learning with such strategies as reflection, goal setting, and self and peer assessment. Reading this report I became increasingly aware that foreign language teachers do not have the tools they require to gather more finely tuned data than the national standards (that is, the assessment opportunities within NCEA) provide. Hattie and Yates (2013) confirmed that, in general, teachers have many tools to measure summative achievement but few tools to measure the developmental steps along the way that allow them to support learning in the manner that the Education Review Office (ERO) recommended.

Motivated by the desire to use data to inform my practice and by the hunch that making time for students to reflect on their language learning would lead to improved performance and, in turn, improved retention rates, I conducted informal research on one year group during two Teaching-as-Inquiry[3] cycles. I asked my students to reflect, first in surveys, then in shared Google Docs, on goal setting and achieving and on the processes and strategies they used for internal and external assessments. I gathered data on their achievement and retention rates and compared them with previous cohorts. My findings indicated superior achievement in national assessments at the ends of Year 11 and 12 for those who had reflected, and an unusually high retention rate from Year 11 into Year 12 and from Year 12 into Year 13. In fact, this retention rate was the best in my subject and in my school over the previous eight years. I wondered if these results were because we had made time to reflect, or whether it was for some other reason, such as a particularly hardworking, talented cohort. More formal research was required.

1.1.2: Four Research Questions

Given these three issues – the diminishing numbers of students opting to learn languages in senior secondary school classes in New Zealand; the elimination of reflection from writing portfolio programmes; and the desire to use data (in addition to my informal inquiries) to inform teaching practice – my doctoral research set out to address the following questions:

1. Does first language written reflection increase second language learning when adolescents write for portfolio assessment? If so, what is the type and amount of increase and for whom?

---

[3] Teaching-as-Inquiry is a professional development cycle recommended by the New Zealand Ministry of Education (MOE) whereby teachers investigate their own practice in order to improve their students' learning outcomes.

2. What insights, if any, does the content of the first language reflections reveal concerning the results of the second language data analysis?

3. Are there any links between the reflective self-regulatory practices reported in the reflections, achievement levels, and/or continued learning?

4. What further insights regarding language learning do the reflections provide?

The terms *first language* and *second language*, as used in the research questions, require further definition. *First language* here equates to English because the research took place in four English-medium state[4] secondary schools. However, it must be acknowledged that, while most participants indicated that English was their first language, more than a third of students (29/71 or 41%) indicated that they also used another language or other languages at home. Thirteen students spoke East-Asian languages (Cantonese, Korean or Mandarin), eight spoke one or more European languages (Bulgarian, Dutch, Flemish, German, Italian, Polish, Serbian or Spanish), four spoke Indic languages (Gujarati, Sinhala, Tamil, or Urdu), three spoke Austronesian languages (Indonesian, Samoan, or Tagalog), and one spoke Indo-Persian languages (Farsi and Pashto).

Second language in the research questions equates to French as Foreign Language (FFL), but could also refer to any additional language, including English as a Second Language (ESL). Applied linguists generally understand *second language* to mean any and all languages that are learned after the first, including the language of the community where the learner is situated as well as foreign language where the community does not speak the language being learned and the learner has comparatively little exposure outside the classroom (Ellis, 2005; Manchón, 2011a; O'Brien, 2004).

1.1.3: Filling a Gap in the Second Language Acquisition and L2 Writing Fields

In addressing these four research questions, this study aimed to fill a research gap given: its topic, the impact of reflective processes on achievement in assessment portfolios; its participants, adolescents in secondary schools learning French; and its area of focus, writing to learn foreign language. To begin with, research on the impact of reflective processes in assessment portfolios is not common. In response to this lack, Belgrad (2013) stated that "[s]tudent reflection on learning in portfolio processes should become a focus of future research. Empirical studies are

---

[4] government funded

needed to test the belief that engagement of students in reflective learning processes has the potential to increase achievement, communication (voice), and motivation" (p. 12). While Belgrad was writing from a primary school mathematics perspective, this study attempted to explore the connection between achievement and reflective processes in assessment portfolios from a secondary school foreign language perspective.

The participants in the study presented here, adolescents learning FFL, are a second important aspect that may make this research interesting to others. Participants in L2 writing research have typically been young adults in universities learning ESL. Although some research has considered younger ESL writers at primary and secondary level, comparatively little has focused on foreign language writing, especially in languages other than English, and especially composed by adolescents (Belcher, 2012; Graham & Macaro, 2007; Manchón, et al., 2009). This study sought to add knowledge to the L2 writing field by researching adolescents in secondary schools learning FFL.

The particular aspect of L2 writing that is the focus of the study, I believe, also renders this research important. Manchón (2011b) called it *Writing to Learn Language* and the *Language Learning Potential of Writing*. This area of research sits at the intersection of the Second Language Acquisition (SLA) and the L2 writing fields and prior to the 1990s was largely unexplored in research literature. This is perhaps because the fields of SLA and L2 writing, despite a common interest in L2 instruction and assessment, have often found each other's research irrelevant. Ortega (2012) called these two fields "unlikely partners (sharing) difficult interdisciplinary dialogues" (p. 404). Whereas SLA researchers have tended to collect experimental evidence from spoken language in order to discover generalisable findings and formulate universal principles of L2 acquisition, L2 writing researchers have focused on what makes for the effective teaching and learning of L2 writing and have tended to conduct more naturalistic research (Ellis & Barkhuizen, 2005). Ortega (2012) suggested that the very specific and comparatively little researched area of foreign language writing could provide an ideal interface where SLA and L2 writing researchers might respectfully inform each other. Schoonen, et al. (2009) described foreign language learning as "an especially interesting setting to study the interaction between higher order skills and lower order linguistic skills" (p. 78) which might lead to a blueprint of the foreign language writer. Perhaps as a result of such sentiments, the field of foreign language writing has experienced a burst of activity since the 1990s and has begun to

develop in its own right, separate from, but informed by, the traditions of both the SLA and L2 writing fields (Reichelt, 1999). It is at this writing-to-learn-foreign-language interface where my research is situated and where it aims to contribute knowledge to the blueprint of the foreign language writer (Schoonen, et al., 2009).

1.2: Setting the Scene

Having introduced the reasons for and the perceived value of my research, this section details the New Zealand context where my research took place and is a response to criticism that L2 research does not adequately address context, thereby diminishing its usefulness to others (Cimasko & Reichelt, 2011; Ellis & Barkhuizen, 2005; Roca de Larios & Murphy, 2001). The following sections detail the New Zealand secondary school foreign language classroom in the hope of making the research more useful to others. They describe the low retention of students into senior language classes, the national foreign languages curriculum, and the national foreign language writing assessment.

1.2.1: Secondary School Foreign Language Classrooms

New Zealand is a member of a group of nations that Kachru (1992) called inner circle English-dominant countries. Given the current hegemony of English, it has to be asked why schools in a geographically isolated, English-dominant nation such as New Zealand, would offer their students additional languages. The Ministry of Education's (MOE) curriculum and resource support website, Te Kete Ipurangi[5] (TKI), offers three reasons why: travel, work opportunities, and personal and national growth and well-being (TKI, 2012).

In response to these three reasons, most students in New Zealand's education system 'taste' a variety of community and international languages in a series of short-term courses while attending primary and/or intermediate school so that, when they arrive at secondary school, aged twelve or thirteen, they are able to choose with a little insight which language or languages they wish to study in earnest. Ironically, although schools are required to offer languages to students in Years 7 to 10 (11 to 14 year olds), the learning of these languages remains optional for the students (East, Shackleford & Spence, 2007). As a result, the number of students learning languages is not high. In July 2015, 245,834 secondary school students were enrolled in L1 English classes (presumably the entire cohort), but only one third (82,817) chose to learn an additional language.

---

[5] The Knowledge Basket

The most popular was our autochthonous language, Maori, chosen by 21,977 students; next in popularity was French with 19,027 students, followed by Spanish with 11,464 students, and Japanese with 10,843 students (MOE, 2015).

Consequently, language teachers are under constant pressure to inspire students to keep learning into the senior secondary school (Years 11 and above, 15+ years of age). In a technology-rich, connected world, it seems logical that foreign language learning would be all the more relevant as target language speakers and their cultures and lands have become all the more accessible. However, the hegemony of English has grown with the internet. A large national survey conducted in England in 2002 revealed that nine out of ten post-sixteen students discontinued foreign language learning. Cohen and Macaro (2007) pointed out that the "[w]orld domination of English has resulted in, motivationally speaking, a losing battle for the study of Foreign Languages such as French and German" (p. 143). A Times Education Supplement article (Garner, 26.8.2016) updated this information:

> The number of (A Level) candidates taking French, still the most popular language, has declined to 9,672 which compares to 31,261 in 1992, and is a 6.4 per cent drop on 2015. The story is the same for German and even Spanish – which has been the only success story in recent years. Even the languages where we are putting most effort into encouraging growth, like Mandarin, are showing a decline, down 8.1 per cent to just 2,849.

A similar pattern is evident in New Zealand. According to statistics published annually by the MOE, of the 40,613 Year 9 students learning an additional language in 2011, four years later, when these students had reached Year 13, only 5,166 remained, that is, 12.7%. This attrition rate is compounded by the drop in numbers beginning to learn a language every year. In 2006, 37% of secondary school students learned another language; six years later, in 2012, only 33% learned a language, a four percent drop; and nine years later, in 2015, 31.5% learned another language, a five and a half percent drop since 2006 (MOE, 2015). East, Shackleford and Spence (2007) attested that "as school rolls have grown, uptake of international languages among secondary school students has declined relative to that growth" (p. 21). Due to small numbers in senior language classes, many schools are forced to combine more than one level per class to keep language learning available in the timetable. This is discouraging for teachers who work tirelessly

to attract and retain students by making language learning relevant and fun and by using the communicative methodologies encouraged by the New Zealand Curriculum (NZC).

1.2.2: NZC *Learning Languages* Curriculum

Since 2007, a requirement in foreign language classrooms in state accredited schools is to follow the *Learning Languages* area of the NZC. Its purpose statement reads: "Students learn to communicate in an additional language, develop their capacity to learn further languages, and explore different world views in relation to their own" (MOE, 2007, p. 17). This three-fold statement leads to the three knowledge strands (or themes) by which the curriculum is organised: communication, language knowledge, and cultural knowledge. It is from these three strands that the proficiency descriptors and achievement objectives for the eight curriculum levels have been developed. These descriptors and objectives state in very broad terms the increasing knowledge, understandings and skills required by students as they progress in their learning. It is up to schools to interpret these descriptors and objectives for their local context.

In their first two years of secondary school, most language students are learning at curriculum levels 1 to 4. By the time they enter Year 11, like the participants in this research, they are embarking on levels 5 and 6. The proficiency descriptor for levels 5 and 6 states: "Students can understand and produce more complex language. They can communicate beyond the immediate context, for example, past and future events. Students can understand and produce a variety of text types" (MOE, 2007, Learning Languages Chart). Developed from this descriptor, the *Communication* objectives state: "In selected linguistic and sociocultural contexts, students will communicate information, ideas, and opinions through different text types, express and respond to personal ideas and opinions, (and) communicate appropriately in different situations" (MOE, 2007, Learning Languages Chart). The *Language Knowledge* objective states: "Students will understand ways in which the target language is organised for different purposes" (MOE, 2007, Learning Languages Chart). The *Cultural Knowledge* objective states: "Students will understand ways in which the target culture is organised for different purpose" (MOE, 2007, Learning Languages Chart). Levels 5 and 6 of the NZC were purposefully aligned with levels A2 and A2+ of the Common European Framework of Reference (CEFR) for Languages (Council of Europe, 2001). With NZC Level 6 language, students are able to meet the NCEA Level 1 examination standards.

The *Learning Languages* area of the NZC actively promotes the CLT methodologies that have been officially promoted in national languages curriculum documents since the 1990s but which have been a component of teacher professional development events and text-book purchases since the 1970s. While communicative methodologies have evolved over this time, the most recent versions still remain true to the guiding principle of communication, and students still learn language through a balance of the four macro-skills, listening, speaking, reading and writing, through activities that focus on meaning as well as accuracy, and through interaction, discovery and analysis. Communicative approaches still view learning as a gradual process of trial and error where learner error indicates the healthy development of an interlanguage (the dynamic, error-laden learner language between the L1 and L2 and the system of rules, faulty or otherwise, on which it is based); and teachers still motivate students with relevant and authentic resources.

In addition to these classical communicative attributes and in response to research, twenty-first century CLT has added new threads to its fabric. More macro-skills have been added (e.g. viewing, presenting and performing); context and learner difference are emphasised more than before; there is more opportunity for collaboration, differentiation and personalisation; there is more scaffolding in the form of strategy instruction; and there is more self-review, peer-review and reflection (Richards, 2006).

This emphasis on communication in foreign language classrooms aligns squarely with the motivations of most New Zealand secondary school language students. In an informal digital survey I conducted in 2013 (n = 244) the top three reasons given by these thirteen to seventeen year olds for learning French related to travel (70%), to the fulfilment of a long-held fascination or dream (52%), or to the desire to become fluent (50%). Because these goals transcend the more immediate goals of enjoyment or examination success, they align with Feuerstein's definition of transcendence which, he theorises, is a prerequisite to learning (Feuerstein & Feuerstein, 1999); Feuerstein believes that, in order for teaching to become learning, students must understand the long-term, transferable purposes of immediate activities (see also Section 1.3.3). For teachers and students alike, therefore, the motivation of foreign language learning is primarily communication. In line with this, "[e]ffective communication is the central aim and the basis for all assessment" (NZQA, 2017, ¶ 8).

1.2.3: Foreign Languages Writing Assessment

In addition to communicative methodologies, the NZC promotes Assessment *for* Learning theory which has been a feature of the national curriculum since 1988 when the direction changing curriculum document, Tomorrow's Schools (Minister of Education, 1988), was published. Assessment *for* Learning encourages learning which results from assessment "by generating feedback that students can act upon in terms of where they are going, how they are going, and where they might go next" (Absolum, et al., 2009, p. 19).

Despite this emphasis on Assessment *for* Learning in the NZC, in the last two decades of the twentieth century, foreign language listening, reading and writing skills continued to be examined by a one off, end of year, timed assessment. The British Assessment Reform Group (2002) called this form of assessment 'assessment *of* learning' because it examines summative language knowledge that is acquired and then retained (East & Scott, 2011). The first step towards Assessment *for* Learning was taken in the mid-1980s with the introduction of an internally assessed speaking component (worth 20% of the final grade), the marks of which were scaled to match the range of a cohort's external examination results. Then in 2002, the NCEA, the current standards-based qualification, was introduced. Not only did it align better with Assessment *for* Learning principles but it also provided an opportunity for language specialists to align assessment with the communicative methodologies also encouraged by the curriculum. Changes included equal weightings for listening, speaking, reading and writing assessments, more classroom-based assessment (about one third), and the development of more appropriate assessment tools (East & Scott, 2011). When the current curriculum, the NZC, was introduced in 2007, the NCEA was once more realigned (2011 to 2013) to match not only the learner-centred expectations of the NZC but also the level descriptors of the CEFR for Languages (NZC, 2007).

The outcome of these recent changes in curriculum and assessment is that foreign language achievement is currently assessed by a balance of Assessment *of* Learning and Assessment *for* Learning approaches (East & Scott, 2011). The reception skills (listening and reading) are assessed entirely via summative end-of-year examination while the production skills (speaking and writing) are measured entirely via internal assessment. A process-based portfolio of different text-types[6] gathered throughout the year is the means by which writing is assessed. As each text

---

[6] Typical NCEA Level 1 text-types include a personal letter, a poster, an email, a blog entry, a personal profile, etc.

is produced, the writing standard (see Appendix A) allows teachers to provide feedback and feed forward so students are able to continue improving their work, using any resources they choose, up until the moment they select the required number of pieces that, in their opinion, best meet the criteria of the standard. These pieces are then submitted on the occasion of a student-teacher conference to be graded holistically by the classroom teacher against criteria that measure language according to text-type, complexity and accuracy and the unhindered communication of ideas (see Appendix B). Grades are awarded according to four broad categories, Not Achieved, Achieved, Achieved with Merit, or Achieved with Excellence, depending on how well the writing meets the criteria of the standard.

These idealistic changes to assessment have had unintended effects on classroom programmes. The frequent changes, the time consuming process of learning and assessing via portfolio, added to the fact that two or even three year groups might be combined into one larger senior class to remain viable, language teachers have often struggled to prepare their students to meet NCEA standards within the time frame of the school year. They have been tempted to take shortcuts by removing unassessed components, such as reflection, from their programmes. In addition, the more time is spent on internal assessment, the less time is spent practising language outside of assessment events and, thereby, the less opportunity there is to receive (in the context of written language production) the direct written corrective feedback that best promotes language learning (Bitchener, 2008). In 2006, the examining authority, the New Zealand Qualifications Authority (NZQA), stated that teachers should give only indirect feedback on writing that would be used for assessment purposes and gave as examples: offering general suggestions for improvement as students write, indicating error by placing an agreed code in the margin of the line where an error occurs, and conducting whole class teaching on common errors (NZQA, 2006). In 2016, the examining authority added that "[t]eacher feedback and feed forward on student's drafts should be holistic to ensure the final product remains a true representation of the student's ability. More than one opportunity for feedback could compromise authenticity" (NZQA, 2016).

Given the pressures that teachers and students were under, in 2016 (a year after the data collection phase of this research took place) the NZQA approved a further change to the assessment criteria for foreign language writing portfolios; they allowed students to select just two pieces of writing to meet the standard instead of the three to five previously specified (NZQA, 2016, February 3). From January 2016 onwards it has therefore been possible for students to

write more unassessed pieces and for teachers to give more direct written corrective feedback on those pieces than was previously possible. (This may or may not be happening in practice.)

While New Zealand's examining body, the NZQA, is to be commended for listening and responding to foreign language teachers' concerns over issues such as workload by making these changes, time pressure in examination years comes from the number of internal assessments across all subjects, not just languages. A longitudinal study conducted by the New Zealand Council for Educational Research based on survey responses indicated that "half the teachers thought that NCEA pressures impacted negatively on student wellbeing, and the NCEA workload for teachers was cited by more than half of principals and teachers as a major issue facing their school" (Wylie & Bonne, 2016).

This means that, while language teachers might have met time pressures by omitting unassessed components from their courses, language students have also acted strategically. Those who continue foreign language learning into the senior secondary school have struggled to balance their dual motivations of examination success and fluency. Students for whom the immediate goal of examination success is more important have been more likely to strategically miss class, not complete unassessed language practice activities, or select standards perceived to be easier to meet (29% of the entire 2013 Year 13 cohort in my own school admitted via a school-wide digital survey to these coping strategies). However, anecdotal evidence from language teachers suggests that those students for whom the transcendent goal of target language fluency is just as important as examination success are more likely to attend class regularly, complete language practice homework that 'doesn't count' for assessment and present work for measurement against all NCEA standards offered. It seems logical that these students would also be more likely to continue with their language learning into the senior school.

It is in the above context that my study took place, a context where foreign languages must be offered for study in the middle school years but where students do not have to choose to learn them, where relatively low retention of students into senior language classes is the norm, where the national curriculum promotes communicative methodologies, and where the national qualification promotes Assessment *for* Learning theory and measures foreign language writing via portfolio assessment.

1.3: Finding a Theoretical Framework and Definitions

Having established the reasons for my research and the context of my research, this section establishes the SLA theory that frames my research. Norris and Ortega (2005) stressed the importance of L2 researchers acknowledging their theoretical base so that careful definitions of research constructs can be provided and associated L2 measurement standards established. In this way even those holding different viewpoints are able to find value in the work (Norris & Ortega, 2005).

1.3.1: Categorising SLA Theories

Selecting a theoretical framework for this research was complex because of the way in which theories of language learning are categorised, the way they draw from other disciplines and vocabularies, the way they build on each other, and the way academics very often use different terms to explain the same or similar concepts or, more confusingly, the same terms to describe different concepts. I was guided in my choice of theory by the work of Arievitch and Haenen (2005), Ellis (2005), Lantolf and Poehner (2014), Norris and Ortega (2005), Polio (2012), and Williams and Burden (1997).

Analysing SLA theory from an educational psychology perspective, Williams and Burden (1997) referred to five schools of thought that have influenced SLA in classroom settings: the behaviourist school which led to audiolingual methodologies, the cognitive and constructivist schools which led to earlier oral-situational and notional-functional versions of CLT, and the schools of humanism and social interactionism which have led to the more recent versions of CLT. Williams and Burden promoted an approach to language teaching that they described as "essentially constructivist... understand[ing] by this that each individual constructs his or her own reality and therefore learns different things in very different ways even when provided with what seem to be very similar learning experiences" (p. 2). However, because learning does not occur in a vacuum, they also sought a social aspect to complement their psychological perspective. They called their approach social constructivism.

Norris and Ortega (2005) divided SLA theory into four approaches according to the era in which they developed and the SLA theories that were prevalent at the time. Nineteen-seventies Generative Linguistics was based on the concepts of Universal Grammar and the theory of learnability. Nineteen-eighties Interactionist Theory added to Generative Linguistics such

concepts as input, interaction, and output. Two more theories developed in the 1990s. Emergentism viewed L2 acquisition as the outcome of neurobiology, while Sociocultural Theory built on Interactionist Theory with the work of psychologists, such as Vygotsky and Feuerstein, who viewed all learning as primarily socially constructed.

Ellis (2005) categorised SLA research into just two approaches, Computational or Sociocultural. Ellis labelled Computational those studies which viewed L2 learning according to psycholinguistic processes and researched such concepts as "input processing, intake, interlanguage development, output processing, and monitoring" (p. 9). Those studies which state that L2 learning originates in social interaction and research such concepts as "the zone of proximal development, private speech, other regulation, self-regulation and scaffolding" (p. 9), Ellis labelled Sociocultural. Ellis (2008) acknowledged that the ten principles of effective instructed SLA that he developed for the New Zealand context were primarily psycholinguistic in nature and therefore Computational. He wrote:

> This model has its limitations and is open to criticism, in particular that it is not socially sensitive because it fails to acknowledge the importance of social context and social relations in the language learning process. It would be clearly useful to attempt to formulate a set of principles based on the broader conceptualisation of second language acquisition – one that emphasizes the importance of the social as well as the cognitive aspects. (p. 5)

While acknowledging the various categorisations made by others before her, Polio (2012) divided SLA theories into six distinct approaches: Generative Theory, Processability Theory, the Usage-based Approach, Skill Acquisition Theory, the Sociocultural Approach, and the Interactionist Perspective. Because her interest was written error correction, she analysed each approach based on its response to four related questions: whether it considered improved written accuracy to be evidence of language acquisition; what role explicit language knowledge played; what role feedback played; and finally, if any written error correction research had been conducted by adhering researchers. Based on her findings, Polio concluded that "the sociocultural approach could be said to be the most unequivocally related to writing" (p. 382). However, she added that, when the essentially speaking-based interactionist perspective is applied to writing, it also has much to add. Polio's suggestion of combining the sociocultural and interactionist aspects of L2

acquisition in one theoretical stance echoes Williams and Burden's (1997) and Ellis' (2008) conclusions.

Lantolf and Poehner (2014) explained why sociocultural and interactionist theories are prime partners; they both stress the importance of meaning in language learning and where one is weak the other is strong – they complement each other. The weakness of Interactionism is that it does not address language development, meaning that research findings can seem inconsistent or contradictory. The weakness of Socioculturalism is that it has to rely on linguistic theory for support. By developing a theory that combines the two, Socioculturalism can provide the developmental theory that Interactionism lacks and Interactionism can provide the linguistic theory that Socioculturalism lacks (Lantolf & Poehner, 2014).

Given the desire to combine cognitive and social features in a theory of language learning capable of both framing my L2 writing research and informing classroom practice, I chose to frame my study within Socioculturalism but also to rely on the psycholinguistic and cognitive concepts of Interactionism. Arievitch and Haenen (2005) claimed that "[l]inking sociocultural and cognitive theories is a necessary prerequisite for developing a framework in which the many now-disconnected findings about teaching, learning, and development can be integrated to meet the growing demands of today's education" (p. 156). The next sections therefore explore the key concepts of sociocultural and cognitive theories that have relevance for this study and, in the process, define the terms used throughout this document.

1.3.2: Interactionist Theory, Language Learning and L2 Writing

Nineteen-eighties Interactionist Theory builds on cognitive linguistics but relates, in particular, to spoken interactions. The theory is concerned with such concepts as interlanguage, comprehensible input, pushed output, attention, noticing and noticing-the-gap, hypothesis-testing, feedback and metalinguistic reflection. This section explains the theory through these terms beginning with the interactionist definition of *learning* and *acquisition*.

In the 1970s, Krashen (1995) distinguished between language that is *acquired* or picked up subconsciously and language that is *learned* by consciously knowing rules. He theorised that there was no interface between them. By contrast, Ellis (2005), in his seminal document that informs the *Learning Languages* area of the NZC, viewed *acquisition* and *learning* as interchangeable terms, preferring instead to differentiate between *naturalistic* and *instructed*

learning and *explicit* and *implicit* knowledge. Ellis described *naturalistic* learning as incidental and taking place in natural settings and *instructed* learning as purposeful and taking place in classroom settings. Learners gain *explicit* knowledge when they pay attention to the language and *implicit* knowledge when they learn unaware. De Keyser (1995) theorised that *explicit* knowledge is converted to *implicit* knowledge through practice, and continued practice leads to fluency.

Ellis (2005) noted the similarity of these concepts to Anderson's Skill-building Theory. Anderson (1993) posited that all knowledge begins in declarative form but becomes procedural through practice. He found that more knowledge is retained when it is used and less knowledge is retained if there is a delay in using it. Feedback and response to feedback play an important role in the skill-building process. When language learning is viewed in the light of Skill-building Theory, declarative knowledge parallels explicit language knowledge, and procedural knowledge parallels implicit language knowledge. Explicit language knowledge becomes implicit through practice and the more language is used the more it is remembered. With continued practice and feedback in new and varied contexts, language becomes automatic or fluent (Polio, 2012).

Another fundamental interactionist term is *interlanguage*. In the early 1970s, Krashen (1995) claimed a natural order to L2 acquisition regardless of input order. Although linguists have since confirmed the existence of developmental sequences, most see too many exceptions to accept Krashen's Natural Order Hypothesis in its entirety. They prefer Selinker's (1972) Interlanguage Hypothesis which notes a loose order to the errors learners make as they progress towards fluency. Selinker called this dynamic, error-ridden language between the L1 and native-like expression an *interlanguage.*

To explain how language is learned, interactionists use such terms as *input*, *output, feedback* and *interaction. Input* describes the language that a learner is exposed to and *output* the language that the learner produces. Krashen (1995) claimed that learners "acquire spoken fluency not by practising talking but by understanding, by listening and reading" (p. 60). Enough comprehensible *input* leads to fluency, and *output* is only important in so far as it elicits more *input*. However, Long (1983) hypothesised that learning occurs when *interactions* are modified to ensure *input* is *comprehensible*. Swain (1985) argued that when learners are encouraged through *feedback* to produce accurate language (*pushed output*) they are more likely to learn it, so, rather than being a result of language learning, output is another step in the learning process. Interaction with others

is therefore an integral component of language learning. Ellis (1994) explained that learning occurs because "*interaction* can provide learners with scaffolding that enables them to produce structures that would be beyond them, if left to their own resources" (p. 26).

Further interactionist concepts related to the language learning process and of particular relevance to L2 writing are *monitoring, noticing*, *hypothesis-testing,* and *metalinguistic reflection. Monitoring* might be described as self-observation, self-judgment and self-reaction (Schunk & Usher, 2013). Closely related to monitoring is *noticing.* Where monitoring is self-initiated, *noticing* (also known as *attention*) occurs as a result of *feedback* from others. As part of her Output Hypothesis, Swain (1985) listed *noticing-triggering*, *hypothesis-testing,* and *metalinguistic reflection* as the three reasons why L2 production leads to L2 learning. Ellis (2005) defined *noticing* as a "cognitive process that involves attending to linguistic form in the input learners receive and the output they produce" (p. 49). This will often involve *noticing-the-gap,* that is, *noticing* the difference between the language produced and native-like language. In her five step sequence in the language learning process, after noticing, Gass (1997) listed linguistic *hypothesis formation and testing*. This involves learners receiving feedback regarding the accuracy of their output and modifying it to render it more accurate. When learners modify their language output based on feedback, this has been "interpreted to indicate that internal hypothesis testing has taken place within the learner's interlanguage system" (Fotos & Hinkel, 2007, p. 133). Metalinguistic reflection, where learners ponder their language use, is an integral part of this input, monitoring, feedback, noticing and output sequence.

In addition to these key interactionist concepts, the terms *form* and *forms,* as used by interactionists, require some explanation. Prior to Long's (1991) definition, these two terms referred to a singular and plural grammar feature. When Long (1991) proposed Focus-on-Form as a more effective teaching approach than either grammar-translation (Focus-on-Forms) or pure CLT (Focus-on-Meaning), the words *form* and *forms* were defined differently. Focus-on-*Form* refers to incidental instruction that causes learners to notice language *forms* in the process of communicative events; Focus-on-*Forms* involves purposeful instruction on pre-selected, isolated language features; and Focus-on-Meaning eschews grammar instruction altogether. Long observed that isolated instruction on grammatical forms failed to promote fluency, while pure communicative instruction failed to promote accuracy. Considering fluency and accuracy to be of equal importance, he conceptualised the Focus-on-Form approach. Through research he

established this to be a more effective methodological alternative to either Focus-on-Forms or Focus-on-Meaning. Focus-on-Form takes the best of both by allowing for the explicit teaching of grammar, but in the context of meaning-centred communicative activities. In his ten principles of Instructed SLA, Ellis (2005) stated that effective teaching focuses on meaning as well as *form* by encouraging learners to notice grammatical *forms* in communicative *input* and to produce those noticed *forms* in communicative *output*.

Polio (2012) examined the relevance of these interactionist constructs to L2 process writing. She equated the learner reading corrections on writing to *input*, rewriting to *output*, and written correction to *feedback* which draws the writer's *attention* to *form.* She found that most *feedback* research in oral contexts was framed by an interactionist perspective and that, in oral contexts, feedback is most effective when implicit and immediate. Because written feedback tends to be explicit and delayed, Polio wondered about the role it plays in writing contexts. She concluded that it is easier to pay *attention* to *form* when writing because the writer has the time needed to do so and has no doubt that corrections have occurred, which is not always the case in oral interactions.

1.3.3: Sociocultural Theory, Language Learning and L2 Writing

Sociocultural Theory is based primarily on the work of psychologists, such as Vygotsky and Feuerstein, who, separately, but from surprisingly similar circumstances, came to complementary conclusions regarding learning (Lantolf & Poehner, 2014). Both Vygotsky and Feuerstein rejected the behaviourist notion of finite intelligence, believing instead in what Feuerstein called alternately *structural cognitive modifiability* and *learning potential* (Lantolf & Poehner, 2014), the infinite ability of the brain to learn new things given appropriate mediation (Feuerstein & Feuerstein,1999). Combined with this belief is the premise that learning takes place via socially mediated occasions which comprise a learner, some tools (both physical e.g. a computer, and symbolic, e.g. language), a more expert other (e.g. a peer or teacher), and carefully designed activities or stimuli, which Bruner (1971) called *scaffolding* (Williams & Burden, 1997).

Feuerstein theorised that learning will only take place when three essential features are present in a *mediated learning event*: *intentionality-reciprocity, transcendence,* and *meaning*. Intentionality-reciprocity is where the teacher is intentional in designing practice activities that are achievable but challenging for the learner while the learner is purposeful in engaging with the learning

activities, responding to feedback and refining output (Feuerstein & Feuerstein, 1999).
Transcendence is when the student is encouraged to understand how the learning extends
beyond the immediate activity to new contexts (see also Sections 1.2.2 and 2.3.2). Equally, when
the learning is imbued with *meaning*, which is also a CLT principle, it is more likely to be
remembered and more easily reapplied.

Vygotsky theorised that learning will only take place when the learning activities sit within the
learner's *Zone of Proximal Development* (ZPD), the area of difference between what learners can
do alone and what they can do with mediation (Lantolf & Beckett, 2009). Learning occurs in this
zone as learners internalise new knowledge with progressively less scaffolding and become
progressively more *self-regulated*. This growing independence related to a specific feature of
learning is known as *microgenesis*, a term first coined by Heinz Werner (1956) in the field of
psychology. Williams and Burden (1997) saw Vygotsky's theory of ZPD as complementary to
interlanguage theory. Because interlanguage is defined as the learner understanding and using
language that is gradually restructured until it more closely fits the target language, the ZPD can
be viewed as "the next level of understanding in the learner's interlanguage" (p. 66).

An aspect of microgenesis is *private or inner speech*, which Knouzi, et al. (2010) described as
self-talk or "speech to ourselves as we regulate our mental functions" (p. 24). Smith (2007) added
that private speech is "not adapted to other listeners, and does not require a response from
others" (p. 341). Because it is private, more often than not it is conducted in the L1 (Ellis &
Shintani, 2014). However, *private speech* changes as the learner becomes more self-regulated. It
begins as external speech where the learner is reliant on the teacher for explicit metacognitive
explanation. As the learner becomes less reliant on the teacher's help, the talk becomes more
strategic and then more implicit until it is purely self-regulated self-talk and often subconscious
(Lantolf & Beckett, 2009). In a language learning context, Knouzi, et al. (2010) associated private
speech with *languaging*, the metalinguistic reflection where learners speak to themselves (or
others) about a language issue.

Language learners are deemed to be fully self-regulated when they use metalinguistic reflection
to fully internalise a language feature (Polio, 2012). Gal'perin (1992) operationalised this passage
towards internalisation in a dynamic four-phase theory of development, testing it in over 800
classes and in many subject areas, including languages (Lantolf & Poehner, 2014). The first

phase involved the teacher creating a non-linguistic model, diagram or picture, a material reminder of the learning to be internalised. In the second phase, this tangible model was replaced by a verbal reminder as the learners carried out the learning activity and explained it to others. The third phase replaced this overt thinking with covert thinking, where learners were encouraged to speak to themselves as they carried out the learning activity. The fourth phase, inner speech, was where the understanding of the learning activity had become an abstraction, fully psychological, automatic and transferable (Lantolf & Poehner, 2014). Sociocultural research in the field of SLA therefore views language as an important cognitive tool to support the learning as well as being the object of the learning.

Another important concept in Sociocultural Theory is Dynamic Assessment (see also Section 2.3.2), which is particularly applicable in L2 writing portfolio contexts. It differs from Assessment *for* Learning in that it is very much a daily teaching and learning partnership unrelated to a specific formative or summative assessment event. Via Dynamic Assessment a teacher ascertains the type of activity a learner requires by measuring just how much control the learner has over the learning and therefore just how much scaffolding can be removed in subsequent learning events. An individual's task performance is consequently never a final indication of ability or even learning potential but an indication of how much mediation is needed in order to move the learning forward (Lantolf & Poehner, 2014).

Budoff (1968), quoted in Lantolf and Poehner (2014), conducted early Dynamic Assessment intervention research by placing participants into three groups according to pre-test scores: high scorers, gainers and non-gainers. High scorers were those who performed well in initial tests with no assistance and who would logically continue to be successful; gainers were those who performed less well in pre-tests but who responded well to mediation; non-gainers were those who failed to improve, even with intervention. From a sociocultural standpoint, lack of progress can be explained as the intervention being outside the learners' ZPD. Where high scorers have already internalised the knowledge, non-gainers require further personalised assistance before progressing further.

Because writing is a socially-mediated tool, Polio (2012) found that, of the six categories of SLA theory that she investigated, the Sociocultural Approach related most naturally to L2 writing. During process writing, in particular, the teacher is able to mediate learning with activities such as

feedback on error, explicit explanation, or other direct or indirect scaffolding according to the needs of the learner. As the learner appropriates the learning and improves the writing, L2 learning occurs (Aljaafreh & Lantolf, 1994).

In light of the above, the sociocultural definition of L2 *learning* is short-term language change which results from support and which occurs in individual events when students require less assistance to complete a task than they needed previously (Polio, 2012). When researching L2 process writing, Swain and Lapkin (1998) considered that the changes that occurred between the original output (first draft) and its reprocessed form (later drafts) constituted learning. In contrast to L2 *learning*, L2 *development* refers to interlanguage change resulting from multiple mediated learning events that occur over time and require increasingly less assistance (Aljaafreh & Lantolf, 1994). L2 *acquisition* refers to the learner's consistent, coherent, accurate and fluent spoken or written language produced without need of support. Such language is internalised and self-regulated (Norris & Ortega, 2005). These are the definitions that I have chosen to work with in this study. I ask whether L1 reflection increases L2 *learning*. I am therefore measuring the short-term language changes that result from feedback and instruction and which occur as a result of the single feedback event that is permitted in the NCEA portfolio writing assessment context.

1.4: Summarising the Chapter and Introducing Those to Follow

Language teachers struggle to cover curriculum and to give students enough unassessed language practice with direct feedback in advance of NCEA portfolio writing. As a consequence, they have tended to omit from their portfolio programmes those elements which are not assessed, such as reflection and self-review. However, if these elements promote language learning, then this omission is short-sighted. My research therefore asked whether L1 written reflection increases L2 learning when adolescents write for portfolio assessment.

In order to ensure that the findings of my research are useful to language teachers and to others in the fields of SLA, L2 writing and beyond, I have endeavoured to describe the research context as thoroughly as possible. To this end, I have explained the trend where fewer and fewer students choose to learn languages in New Zealand's senior secondary school and have detailed the broader New Zealand classroom context in which my research is situated: the NZC which promotes communicative and Assessment *for* Learning approaches, and the high-stakes national qualification, NCEA, which assesses foreign language writing via portfolio assessment.

I have also shared my search for an appropriate theoretical framework for my L2 writing research, deciding (with many others) on the suitability in L2 writing portfolio contexts of Sociocultural Theory combined with psycholinguistic elements of Cognitive-Interactionist Theory. Summarising both theories provided an opportunity to define the terms to be used throughout this document; of particular importance is the chosen definition for L2 learning as used in the research questions: the short-term language changes that result from feedback and instruction, which occur as a result of the single feedback event that is permitted in the NCEA portfolio writing assessment context, and which indicate interlanguage change and movement towards self-regulation.

Chapter Two sets out to review L2 writing literature and literature related to reflective learning practices. The review begins by establishing that writing has the potential to promote language learning when learners are given the time and the guidance to plan, compose and review as they write, and to notice linguistic forms and reflect in such a way that deep, explicit, metalinguistic processing occurs (Manchón, 2011a). The NCEA writing portfolio is an ideal assessment format for such reflective language learning. As a result of the literature review, learning habits, goals, metalinguistic reflection on teacher feedback and language learner strategies were established as suitable topics for reflection, and the learners' first language and online blogs were established as suitable modes for reflection.

Throughout the literature review, it was difficult not to notice the number of calls by researchers for more investigations by classroom practitioners. Chapter Three reports on a pilot study I conducted as a practitioner-researcher. I wanted to locate and trial objective text-analysis tools suitable for measuring L2 learning in the larger study to follow and which might also be suitable for classroom teachers wishing to find more fine-grained measures of language learning to use in conjunction with the more holistic rubrics typically used in portfolio assessment contexts. My pilot study found that error analysis and situated complexity, accuracy, and fluency measures were the most suitable of the five tools trialled for measuring foreign language learning in secondary schools.

Chapter Four builds on the pilot study and looks at research methodology literature to find a research design suitable for measuring the impact of L1 reflection on L2 learning. Because a theoretical framework that combines sociocultural and interactionist theories of language learning was chosen to frame the study, a mixed method approach was deemed the most suitable for the

research design. Utilising a mixed method, counterbalanced, comparative study which ensured no student was advantaged or disadvantaged in the high-stakes assessment context in which it took place, the quantitative L2 data collected from the participants' writing was supplemented by the L1 qualitative data collected from their written report. A quadrant of measures capable of tracking small, incremental and daily steps in interlanguage development, both in L2 writing data and in L1 reflection data, is presented in this chapter.

After setting out the background to the research in the first four chapters, Chapter Five reports on the first half of the study. It outlines the four investigations that measured the impact of the reflection intervention on the L2 written proficiency, complexity, accuracy and fluency, verb error elimination, and verb self-regulation of intervention and comparison groups and NCEA subgroups. The studies found that, while the reflection intervention had little impact on verb error elimination or verb self-regulation, it did have a statistically significant impact on the L2 written proficiency and on the accuracy of the Term 1 Merit intervention group in particular.

Chapter Six details the second half of the study. Based on the links between SLA and self-regulation established by Teng and Zhang (2016), this chapter analyses the L1 reflective blogs that accompanied the L2 writing through the lens of Zimmerman's (2013) dimensions of self-regulated learning. It aimed to find insights into why the Term 1 Merit participants gained from the reflection intervention when other groups did not. It found that the L1 language and the depth of thinking evident in the blogs, the strategies employed while L2 writing, the goals set for improving L2 writing, and the goals related to continuing to learn all approximated the participants' NCEA grades. It also found that the Term 1 Merit participants, whose L2 gain was statistically significant, were at the lower end of the Merit band and had more potential for improvement than their Term 2 counterparts.

Chapter Seven discusses, synthesises and interprets the results of the four investigations in Chapter Five and the findings from the six dimensions of self-regulated learning in Chapter Six. It offers three interpretations for the Term 1 Merit participants' statistically significant gain in learning: one from a sociocultural perspective, one from a skill-building perspective, and one from a metaphorical perspective. From a sociocultural perspective, it may be asserted that the Term 1 Merit participants who benefitted most from the reflection intervention were at a stage in their ZPD and interlanguage development where they were ready to benefit in terms of language learning

from the mediation of the reflection intervention. They were the 'gainers' described by Budoff (1968). Those who did not appear to benefit were either the Excellence participants, who were already 'high scorers' at the self-regulated end of the microgenetic scale, or the Achieved participants, the 'non-gainers', who needed more explicit scaffolding, in addition to the mediation of reflection, before being able to move their learning forward (Aljaafreh & Lantolf, 1994). From a skill-building perspective, those who benefitted from the reflection intervention were those who were able to use the reflection to respond to the indirect feedback offered to improve their writing. This equated to more practice which in turn moved the learner's declarative knowledge closer to procedural knowledge and, ultimately, fluency. From the metaphorical perspective of a 'bi-cycle', the impetus that propelled the language learners along their ZPD-Interlanguage Development Road when L2 writing was the self-regulation cycle of self-observation, self-judgment and self-reaction (Schunk & Usher, 2013) and the language acquisition cycle of noticing-triggering, hypothesis-testing and metalinguistic reflection (Swain, 2006).

Chapter Eight is the concluding chapter which looks back at the study and looks ahead to future studies. It revisits the three issues that prompted the study in the first place – the declining numbers of students in New Zealand's senior secondary school language classes, the omission of reflective practices in assessment portfolio programmes, and the lack of measuring tools that are able to collect the kind of data that informs daily teaching and learning in secondary school foreign language classrooms. All are addressed to a certain degree by this doctoral study. As explained in Chapter Four, a quadrant of measures capable of tracking small incremental changes in L2 learning and in L1 reflection data was developed for use in the study. The study investigated the impact of omitting reflective practices from writing portfolio assessment. In the single mediation event that formed the reflection intervention, it found that only those students who were already in the ZPD benefited from the reflection intervention. It raises the possibility that, if this study had offered further opportunities for reflection and over a longer period of time, then students from all three groups, Achieved, Merit and Excellence, might have benefitted at different times in the learning programme and related to different language features. The study concludes by encouraging language teachers to begin the academic year by making time in their busy writing portfolio programmes for students to develop a habit of reflecting on the processes and strategies they use when writing and to set goals that relate directly to deficits in learning. If practised regularly throughout the academic year, reflective practices might be instrumental in

moving their learning forward and in encouraging them, through increased understanding,

motivation and self-efficacy, to continue learning in subsequent years.

CHAPTER TWO: REVIEWING THE LITERATURE

In the previous chapter I introduced the reasons for my research, the context in which my research took place, and the theory by which my research is framed. In this second chapter I survey research literature to discover what has already been said about L2 writing so that my study builds on the work already produced by others. The chapter is divided into four sections that survey research literature on the categories of L2 writing research, the language learning potential of L2 writing, the measurement of L2 writing, and reflecting when L2 writing.

2.1: Surveying L2 Writing Research

Because L2 writing is both a cognitive process and a situated activity, L2 writing research includes a broad array of topics that emanate from multiple research traditions and from theoretical diversity (Cumming, 2016; Manchón, 2016). Polio and Friedman (2017) explained that this complexity and diversity is the result of having two parent fields, the cognitively oriented, quantitative, and experimental research field of Second Language Acquisition (SLA), and the socially oriented, qualitative, and naturalistic research field of L1 composition. While these two fields represent two different research paradigms, postpositivist and postmodern, Polio and Friedman posited that they are not contradictory but complementary and result in the L2 writing field's openness to "a range of research paradigms, approaches, and methods that has allowed for the emergence of a body of work that has enriched understandings of the complex social and cognitive processes that constitute L2 writing" (Polio and Friedman, 2017, p. 4).

In order to understand a little of the scope of L2 writing research, its methodology and findings, this section surveys literature that categorises the research (see Table 1). In the first decade of the twenty-first century, the categories seemed to be similar and tended to align with Archibald and Jeffery (2000), who grouped L2 writing research into four main categories according to the focus of the research, whether that was on process, product, context, or pedagogy. For example, Cumming (2001) grouped L2 writing studies into three main categories according to whether they analysed the features of the texts produced, the processes used while writing, or the sociocultural contexts in which the writing took place. Ransdell and Barbier (2002) also grouped L2 writing studies into three main categories but according to their linguistic, psycholinguistic, or pedagogical emphases. Zhang (2008) grouped research into five major areas of inquiry, whether on L2 writers' characteristics, L2 writing process, instruction and feedback on L2 writing, or on the

L2 writers' texts. Hyland (2009) grouped L2 writing research according to the practitioners, the writers' texts, the writers' processes and attitudes, and the readers' expectations and evaluations.

Table 1: Categories of L2 Writing Research

Date	Author	Categories			
2000	Archibald & Jeffery	Product	Process	Context	Pedagogy
2001	Cumming	Features of texts produced by writing	Processes used while writing	Sociocultural contexts of writing	
2002	Ransdell & Barbier	Linguistic	Psycholinguistic		Pedagogical
2008	Zhang	Writers' texts	Writing process and feedback on writing	L2 Writers' characteristics	L2 writing instructions
2009	Hyland	Writers' texts	Writers' processes and writers' attitudes	Readers' expectations and evaluations	Practitioners
2016	Cumming	Contrastive rhetoric	Cognitive models of composing	Genre theories	Sociocultural theory
2017	Polio & Friedman			Writing for real life purposes	Writing to learn language
2011a	Manchón	Descriptive		Interventionist	
		Cognitive Focus on process, Focus-on-Form, input, output, noticing	Sociocultural Focus on interaction or collaboration	Cognitive Focus on process, Focus-on-Form, input, output, noticing	Sociocultural Focus on interaction or collaboration

In the second decade of the twenty-first century, the categorisation of L2 writing research appeared to be more cognisant of researchers' theoretical beliefs and their subsequent choices of methodology as well as the focus of the particular study. For example, Cumming (2016) added to his previous groupings, four primary theoretical stances: contrastive rhetoric, cognitive models of composing, genre theories and sociocultural theory. He also referenced critical theory, goal theory and dynamic systems theory. By contrast, Polio and Friedman (2017) indicated more simply that L2 writing research fitted two basic types according to pedagogical purpose, whether the writing was to learn language, or whether it was for real life purposes.

However, the categorisations of Manchón (2011a) are those that give direction to the next two sections. Manchón placed L2 writing research first of all into categories according to the methodologies used, whether descriptive or interventionist, and then divided both types according to the theoretical frameworks that informed them, either cognitive or sociocultural. Within these groupings she then placed the projects, as previous meta-surveys had done, according to focus. The next section details examples of L2 writing research using Manchón's descriptive or interventionist, and cognitive or sociocultural categories.

2.1.1: Eight Descriptive Studies of L2 Writing

Descriptive research explores the chosen focus through observation and analysis. This section summarises four studies that fit Manchón's (2011a) descriptive cognitive category (Roca de Larios, et al., 2008; Ruiz-Funes, 2015; Sasaki, 2000; Torras & Celaya, 2001), and four studies that fit her descriptive sociocultural category (Aljaafreh & Lantolf, 1994; De Guerrero & Villamil, 2000; Laoire, 2014; Teng & Zhang, 2016).

Torras and Celaya (2001) described and analysed the writing of two groups of English as Foreign Language (EFL) learners who had started learning at different ages, eight and eleven years old respectively (n = 63). Using a longitudinal research design, three pieces of writing were measured in intergroup and intragroup analyses. Torras and Celaya found that the students' text proficiency improved consistently over time according to complexity, accuracy and fluency (CAF) measures but that these three aspects of language developed at different rates. They also found that those who began language learning younger did not show significant advantages in complexity, accuracy or fluency when writing over those who began learning later.

Roca de Larios, et al. (2008) investigated the time allocated to different foreign language process writing activities by three groups of students who were at three different proficiency levels in a Spanish secondary school and university (n = 21). The researchers collected data using think aloud protocols while the participants wrote an argumentative essay in a one hour period. They found that most of the time (60-80%) went to text generation. Roca de Larios, et al. also found that the least proficient writers (secondary school students with six years of EFL learning) spent more time formulating (converting ideas into language) and less time critically evaluating and revising their writing than more proficient writers (graduates with twelve years of EFL learning). They deemed this important because evaluation and revision are the sorts of written output activities that contribute to L2 learning.

Sasaki (2000) described the writing processes used by EFL learners with different levels of L2 proficiency (n = 12). Using stimulated recall protocols, she analysed students' explanations of their writing behaviours as they watched themselves on video immediately after writing. Sasaki noted the changes that took place in the students' writing processes over time and found that differing amounts of time were spent on planning, editing and translating from L1 to L2 according to the prior L2 experience and also the ability of the writer. As a result, Sasaki was able to

differentiate her curriculum delivery according to three student levels: expert, novice before instruction, and novice after instruction.

Ruiz-Funes (2015) explored the relationship between the language learning potential of L2 writing and task complexity. Her participants were 32 Spanish as Foreign Language learners from advanced and intermediate courses in a US university. Each participant wrote two tasks, one more cognitively challenging than the other, and was awarded one of three performance levels, high, medium or low. Ruiz-Funes scored the two pieces of writing using multiple CAF measures. Although not statistically significant, she found a tendency in both the advanced and intermediate learners for increased complexity but decreased accuracy and fluency when responding to the more challenging task, and decreased complexity but increased accuracy and fluency when responding to the less challenging task. Ruiz-Funes interpreted these findings in light of working memory capacity and a trade-off in attention when L2 composing. However, when investigating further according to the three performance groups, she found that the high performing learners in the advanced learner group showed a greater increase in complexity as well as accuracy and fluency when responding to the more complex task than when responding to the less complex task. It appeared that the high proficiency, advanced learners were inspired by the more complex task and were able to pay attention to all aspects of language. Ruiz-Funes concluded that, for these learners, task complexity increased the language learning potential of L2 writing.

Where the four examples above were categorised as descriptive studies with a cognitive focus, the next four are descriptive with a sociocultural focus. Aljaafreh and Lantolf (1994) asked whether negative feedback on error in L2 writing leads to L2 learning. They followed nine adult ESL volunteers as they responded via L1 conversation to feedback on their L2 writing and analysed their responses according to five microgenetic phases, from total dependence on the teacher to full self-regulation, and from no knowledge of a language feature to full appropriation of that feature. They concluded that "[e]ffective error correction and language learning depend crucially on mediation by other individuals, who in consort with the learner dialogically co-construct a zone of proximal development in which feedback as regulation becomes relevant and can therefore be appropriated by learners to modify their interlanguage systems" (p. 480). From their perspective, all types of feedback become potentially relevant for L2 development.

De Guerrero and Villamil's (2000) descriptive, sociocultural study also used a microgenetic approach to analyse the revision strategies of two intermediate level ESL students, one a reader and the other a writer. The researchers wanted to "observe the mechanisms by which strategies of revision take shape and develop in the interpsychological space created when two learners are working in their respective ZPDs" (p. 51). De Guerrero and Villamil analysed the scaffolding mechanisms used by the reader and the writer as they collaboratively revised the document written by the writer. They coded the assistance mechanisms using a combination of categories drawn from previous studies, which included recruiting and maintaining interest, noting inconsistencies in the text, teaching each other, and modelling examples of better language. The researchers found that, in the process of improving the document in this interpsychological space, the students produced mutual rather than unidirectional scaffolding.

Laoire's (2014) descriptive, sociocultural research used L1 narrative to investigate the extent to which L2 learners of Irish (n = 15) in five Irish secondary schools were able to transfer skills, strategies and experiences from their compulsory L2 Irish learning to their optional third language (L3) French, German or Spanish learning. The participants wrote two L1 English extracts about their L2 and L3 language learning and then reflected on what they had written in a semi-structured interview. Laoire found that all learners viewed being 'good' or 'not good' at Irish as something they could not alter, and that they dealt with any difficulties in their Irish learning by rote learning and memorising chunks. Interestingly, three learners (20%) indicated that they were better at their more recently learned L3 than they were at their long-standing L2. However, only three (20%) were able to make crosslinguistic references between their L2 and L3. Laoire described the learners' metalinguistic knowledge as superficial and lacking in reflection and engagement. She concluded that the unhelpful tradition of teaching Irish via a direct method that sought to avoid contamination from English was part of the reason. She called for language awareness programmes to help learners capitalise on their L2 Irish learning for the benefit of subsequent language learning.

Teng and Zhang (2016) conducted descriptive, sociocultural research in six North-East Chinese universities with 790 undergraduate EFL writers. Because "[i]t appears that self-regulation is pivotal to student success in L2 learning in academic settings" (p. 674), the researchers sought empirical evidence for using self-regulation theory in the field of L2 writing by validating a newly-developed, self-report instrument, *The Writing Strategies for Self-Regulated Learning*

Questionnaire. Using confirmatory factor analysis and structural equation modelling, they compared three structural models to represent the dimensions of EFL writing strategies for self-regulated learning, and they asked whether EFL writing strategies for self-regulated learning are able to predict EFL writing proficiency. They found that four dimensions of self-regulation (cognition, metacognition, social behaviour and motivational regulation) correlate with nine frequently used writing strategies, and that six of the nine strategies had significant predictive effects for EFL writing proficiency. They recommended the newly-developed questionnaire as an effective self-evaluation tool for students and a way for teachers to increase their students' awareness of self-regulated writing strategies and, given the correlation between the strategies and L2 proficiency, a way for teachers to increase learning outcomes.

2.1.2: Eight Interventionist Studies of L2 Writing

Interventionist research on L2 writing begins by establishing the context in which the writing takes place (a pre-test) and offers an intervening input, such as feedback, exemplars to analyse, or time to reflect. It then analyses language production (a post-test) to establish the impact of the intervention. L2 studies that fit this category include the cognitive studies by Adams (2003), Bitchener (2008), Izumi (2002), and Ong (2014), and the sociocultural studies by De Silva and Graham (2015), Knouzi, et al. (2010), Lamy and Hassan (2003), and Lee (2015).

Izumi (2002) conducted an intervention that researched the cognitive roles of input and output in the development of English relative clauses. The participants were adult intermediate level ESL students who were enrolled in several US universities (n = 61). They were placed into four treatment groups and one comparison group and were asked to read and understand a text, to reconstruct the text (if they were in one of the two output groups) or to answer questions on the text (if they were in one of the two non-output groups). Izumi wanted to discover whether the act of producing output caused learners to notice the target form to a greater extent or whether enhanced input aided the noticing of the form better. Izumi found a three-fold greater learning gain in the output group and little gain in the enhanced input group. He argued that "output triggered deeper and more elaborate processing of the form which led them to establish a more durable memory trace" (p. 570).

Adams (2003) replicated and extended Swain and Lapkin's (1998) feedback research with a cognitive intervention study that collected data from the audio recordings of 23 pairs of L2

Spanish learners as they reconstructed a story from pictures and then collaboratively wrote the story. L1 speakers then reformulated each piece of writing. The intervention involved randomly assigning the pairs into three groups: a control group who simply repeated the task, a noticing group who repeated the task and then compared it to the L1 speaker's reformulation, and a noticing plus stimulated recall group, who repeated the task, compared it with the L1 speaker's reformulation and then, by stopping and starting the recording of their collaborative rewriting, recalled their thoughts at the time they made the comparisons. Finally, a week later, each individual rewrote the story based on pictures alone. Adams found that learners in the stimulated recall group used significantly more new language forms in their post-intervention output.

In a two month cognitive intervention, Bitchener (2008) researched four groups of low intermediate ESL students (n = 75) as they responded to feedback on the use of the referential definite and indefinite articles, 'a' and 'the', in three pieces of writing produced over two months, a pre-test, an immediate post-test, and a delayed post-test. The first group received direct corrective feedback and written and oral metalinguistic explanation, the second group received direct corrective feedback and written metalinguistic explanation, the third received direct corrective feedback only, and the fourth (the control group) received no corrective feedback. Bitchener found that "the accuracy of students who received written corrective feedback in the immediate post-test outperformed those in the control group and that this level of performance was retained two months later" (p. 102).

Ong (2014) conducted a cognitive intervention by examining the effects of four different planning and writing times and three different task conditions on the use of five metacognitive processes: generating new ideas, elaborating new ideas, organising new ideas, thinking of essay structure, and thinking about language aspects of the task. The106 pre-university L2 learners in a Singaporean EFL programme wrote a thirty minute argumentative essay under different conditions. One group planned for 10 minutes and wrote for twenty, a second group planned for twenty minutes and wrote for ten, a third group had no planning time but wrote freely for 30 minutes, and the control group received no instructions on the use of their time. The task conditions were changed by varying the essay prompts. The control group received just the question, a second group received the question with ideas to include, and a third group received the question, the ideas and a macro-structure for setting out the essay. The participants completed a strategy use questionnaire immediately after writing because the researcher believed

that this was the least intrusive method when researching with a large number of participants and a complex study design. Ong found that writers who received just the question reported generating and organising more new ideas in the planning stage and elaborating and organising more new ideas in the writing stage than others who had more support. She also found that writers who were given planning time engaged in significantly more planning during the writing stage than those who were not given planning time. Interestingly, the planning groups reported a significantly higher amount of thinking about language than the other groups.

Where the four previous studies were cognitive interventions, the next are sociocultural interventions. Lamy and Hassan (2003) conducted sociocultural intervention research with adult FFL learners (n = 4) in a longitudinal (15 months) study that focused on the relationship between learner behaviour and task design – unstructured, semi-structured and highly structured. The researchers studied whether reflective interaction was likely to occur more during some learning tasks than others. They gathered data from a content analysis of the learners' online messages and from questionnaires on feedback. They concluded that task-type is not the main initiator of reflective interaction. Instead, they found that, regardless of the task, the students adopted one of two levels of language learning, either deep or surface. They defined surface level learning as manipulating and memorising lexical items and deep level learning as that which requires a more critical understanding of the lexical items and contains a reflective element.

Knouzi, et al. (2010) explored self-scaffolding mediated by 'languaging' in a sociocultural intervention study. The researchers understood from previous research that those students who 'language' more (that is, who self-scaffold their learning by talking about language problems) achieve a higher level of accuracy and a deeper level of linguistic understanding. They also understood that those students who use only one type of languaging do not perform as well as those who use a balance of languaging types. Based on these understandings, they conducted a multi-stage, microgenetic analysis of the languaging used by two university FFL students as they read about and talked through the concept of voice. The researchers found that "languaging mediates the conceptual development of language learners and facilitates the internalisation of declarative knowledge" (p. 45). They also found that high 'languagers' were able to make better connections between new pieces of knowledge and prior knowledge. When they encountered cognitive conflict, they languaged and hypothesised about it, and then confirmed or rejected their

hypotheses. Knouzi, et al. called for "a change in educational practices that would allow for more learner agency through self-scaffolding mediated by language" (p. 23).

In a sociocultural intervention, Lee (2015) investigated the opinions of thirty junior secondary school students in Hong Kong regarding feedback on their L2 writing. During a two-week research period, the participants wrote a two page essay, took part in a one hour feedback training session, then, in randomised groups of three or four, participated in different types of feedback exercises before rewriting their two page essays and completing a questionnaire and an audio-taped, semi-structured interview. The types of feedback included reading and commenting on peers' essays, discussing one's own feedback to a peer with a partner (intra-feedback), reading peer comments on one's own writing, reading teacher feedback on one's own writing, and discussing peer comments in an oral response session. Lee found that participants preferred teacher feedback to peer feedback because of the teacher's expertise and the quality of the feedback offered. However, they still valued taking part in peer feedback activities because the sessions were constructive and fun, and they made the participants more aware of and attentive to the readability of their writing. In short, peer feedback was motivating. The participants also indicated that working with a peer to give feedback gave them more confidence in their judgments. Lee concluded that peer feedback was a valuable inclusion in junior secondary school programmes and that the benefits of intra-feedback, in particular, require further investigation.

De Silva and Graham (2015) conducted a sociocultural intervention with 12 undergraduate students in a Sri Lankan English for Academic Purposes class to discover the impact of writing strategy instruction on strategy use by high and low attainment students. They defined writing strategy as "a writer's conscious mental activity employed in pursuit of a goal (i.e. in order to solve a problem) within a particular learning situation, and which is transferable to other situations and tasks" (p. 49). Over a twenty-four week course, the experimental group received strategy instruction in addition to L2 writing instruction while the control group did not. The strategies taught were those that a pre-test indicated were not used well by the participants, namely, planning, monitoring and editing strategies, and, in addition, the metacognitive strategies related to task analysis. From questionnaires, diaries and stimulated recall, De Silva and Graham found that, for both high and low achievers, strategy instruction resulted in greater adeptness of strategy use as evidenced in the combining of strategies by intervention groups, and that this resulted in improved writing, more metacognitive awareness and a greater sense of direction. They also

found that the use of stimulated recall immediately after L2 writing, rather than other data collection methods, strengthened the findings because it was not intrusive on the writing process of the learner. De Silva and Graham called for more research on the effectiveness of stimulated recall for strategy research.

In the two sections above, sixteen L2 writing studies were categorised according to Manchón (2011a) in order to establish the scope of research that has already been conducted in the field and in order to build on what has gone before. All sixteen studies have contributed in some way to the current research (see Section 2.5). In addition, all have emphasised the language learning potential of L2 writing which is the subject of the next section.

2.2: The Language Learning Potential of L2 Writing

In English-dominant nations, Reichelt (2011) claimed that there is little inherent value in foreign language writing because there is little need for it, its value deriving from its ability to promote language learning in classroom contexts. Indeed, foreign language teachers have long assumed the language learning role of writing (Benevento & Storch, 2011). For them, writing is one of the four CLT macro-skills alongside listening, speaking and reading; and each macro-skill has two primary functions in their classrooms: one formative, to learn language, and one summative, to provide evidence of language knowledge (Reichelt, 1999).

Researchers, on the other hand, while they might consider speaking and writing to be more reliable than listening and reading to indicate what language learners know (Ellis & Barkhuizen, 2005), they have expressed mixed opinions on "the instrumental role that writing can play in the acquisition of a second language in educational settings" (Harklau, 2002, p. 345). Some, like Wolff (2000), insisted that writing is probably the most important L2 learning tool available. I. Lee (2011) agreed that writing undoubtedly plays a very important role in a foreign language classroom. Some would go so far as to say that writing is "a central mechanism via which language competencies, especially advanced competencies, may be, or perhaps must be acquired" (Norris & Manchón, 2012, p. 221). Others, like O'Brien (2004), cautioned that these claims have not yet been substantiated by empirical data. However, Manchón (2011a) insisted that there is ample empirical evidence to indicate that writing contributes to language learning.

She supported her belief in the *Language Learning Potential*[7] of writing by referring to four evidence-based theories, the Noticing and Output Hypotheses, Focus-on-Form and Skill-building Theory. In the following sections, these four theories are paired with research studies that provide empirical evidence indicating that, when L2 learners write, they create an environment in which language learning is able to occur.

2.2.1: Noticing and Output Hypotheses

The first two theories that Manchón (2011a) promoted in support of her belief in the language learning potential of writing were the Noticing and Output Hypotheses (see also Section 1.3.2). These assert that language learning occurs when learners are prompted to produce *output* and then modify that *output* by *noticing* the difference between the language they have used and L1 speaker language. Swain and Lapkin (1995) investigated thirteen-year-old immersion students' responses to noticing a gap between what they wanted to write and what they were able to write. Using think aloud protocols, the researchers found that, when participants were given sufficient opportunity to modify their output based on the feedback received, and when they reflected on their modifications, the cognitive processes integral to L2 acquisition were activated. They concluded that "[w]hat goes on between the original output and its reprocessed form … is part of the process of second language learning" (p. 371). They listed these cognitive processes as noticing-triggering, hypothesis-testing, and metalinguistic reflection.

In his MOE-commissioned survey of research literature, Ellis (2005) summarised multiple studies that explained why written output is essential to L2 learning: it elicits targeted and individualised responses from the teacher; it causes the learner to notice syntax; it gives opportunity for linguistic hypothesis-testing and corrective feedback; it moves the learner closer to fluency through practice; it provides opportunity for output that is longer than a paragraph which develops discourse skills; it allows for personal voice; and it provides opportunity for students to become independent learners as they reflect on their writing, which, Ellis stated, is equivalent to auto-input.

---

[7] It is interesting to compare Manchón's use of the term *Language Learning Potential* with Feuerstein's use of the term *Learning Potential*. Was Manchón influenced by Feuerstein or did she devise the term independently? Either way, both terms are pertinent to the current research.

2.2.2: Focus-on-Form Theory

The third theory that Manchón (2011a) used to support her belief in the language learning potential of writing was Long's (1991) Focus-on-Form (see Section 1.3.2) which encourages learners to notice and reproduce grammatical forms in the course of communicative activities. Harley (1989) observed that, after many hours of immersion, students in a Toronto French immersion school were still struggling with the uses of the perfect and imperfect tenses. She conducted an intervention study with 319 Grade 6 students in twelve immersion schools, six schools as intervention groups and six as control groups. Harley found that, after direct instruction, the usage of those in the intervention groups improved faster but that the difference was not maintained three months later. On revisiting her data and interviewing the teachers, Harley found that one of the teachers had focused on the forms more than the others and that her class had outperformed all other classes. Harley (2007) stated that, "once a second language feature has been identified as problematic, there is merit in preplanning a segment of the curriculum to deal with it" (p. 171). She advised teachers that complex structures require more form-focused instruction because the more complex the language structure the more unlikely it is to be noticed and correctly reproduced.

Fotos and Hinkel (2007) also concluded that meaningful input alone is insufficient to develop L1-like accuracy in writing, regardless of the language level of the participants. Input must be partnered with output that includes reproducing newly instructed forms in communicative activities. However, they reminded teachers that Focus-on-Form does not necessarily mean explicit rule teaching and that communicative options are also available. They quoted research previously conducted by Fotos in three semester-long EFL classes where the students were participating in an email exchange with L1 speakers. The teacher taught only the forms that the students needed based on the common errors they had committed during the previous week's emails. The teacher then found online examples of correct usage of the forms to display on the students' screens. In this way the students encountered communicative, real-world examples immediately after the form-focused instruction which was based on their most recent output errors. Based on their research, Fotos and Hinkel recommended a form-focused writing curriculum that includes pre-writing planning, explicit form-focused instruction, ample opportunities for output, corrective feedback, reactive form-focused instruction, and multiple learner revisions.

Ellis (2005) included Focus-on-Meaning as well as Focus-on-Form in his list of principles for effective instructed SLA. He suggested that teachers encourage learners to notice forms in input and give learners ample opportunity to produce output containing the noticed forms. He advised teachers that corrective feedback should follow. By implementing this sequence, teachers create rich environments for language learning.

2.2.3: Skill-building Theory and the Controversial Role of Feedback

The fourth theory that Manchón (2011a) quoted to support her belief in the language learning potential of writing was Skill-building Theory, the theory that declarative (explicit) knowledge becomes procedural (implicit) knowledge as a result of repeated practice (see Sections 1.3.2 and 7.3). An essential element in Skill-building Theory is feedback and the learner's response to feedback in repeated practice opportunities (Anderson & Corbett, 1993).

Anderson and Corbett (1993) wanted to know if Skill-building Theory had a practical application in everyday classrooms so they built and researched a computerised tutoring system that modelled a new mathematical skill as a set of production rules, and which taught those rules through a series of examples. The virtual tutor gave students immediate but minimal feedback so that they might deduce the rules from the examples as quickly as possible and control the amount of explanation they received. In this way, it was hoped that knowledge retention would be enhanced and independent learning skills developed. However, the researchers found that the students failed to seek enough explanation from the virtual tutor, so they intended to further their research with different methods to motivate students to read and process explanatory messages more deeply and so increase their learning rate.

Anderson and Corbett's (1993) Skill-building research conducted in mathematics classrooms aligned with Bitchener (2008) who conducted research in ESL classrooms and found that students who were given immediate written corrective feedback in the form of metalinguistic explanations outperformed others and that the learning was still intact two months later. It also aligned with Ellis (2005) who stated that when students are allowed to provide their own input (auto-input) on their output, independent learning is promoted. From a Skill-building perspective, language learning results because immediate feedback optimises learning rates and auto-input allows students to generate their own understandings of language rules which, in turn, enhances

knowledge retention; and written metalinguistic explanations allow students to process explanations on error more deeply and thereby increase learning rates (Anderson, 1993).

However, the effectiveness of feedback on language error, regardless of repeated practice, is a controversial issue. From his Generative Linguistics stance, Truscott (1996) began the controversy when he found that written corrective feedback does not lead to L2 acquisition because, when learners focus on accuracy, other aspects of their language, such as complexity, suffer, and because corrective feedback builds explicit knowledge, not the implicit knowledge that leads to L2 acquisition.

From their respective studies into the profile of the 'good language learner', Roberts and Griffiths (2008) concluded differently. They found that good language learners notice and understand correction and use corrective error feedback to improve their language abilities, whereas poorer language learners tend to ignore correction.

From a cognitive-interactionist standpoint, Ferris (2006) analysed the responses to teacher feedback on four scripts written by 93 adult ESL immigrants in an undergraduate composition class over the period of one semester. She found a strong relationship between teacher error marking and successful student revisions (or practice opportunities if seen from a Skill-building perspective) and that this progress in accuracy lasted over time. Ferris also found that 89% of teacher feedback was accurate and complete, that students corrected errors in a wide range of linguistic categories, and that, in the context, indirect feedback was superior to direct feedback.

Also from an interactionist viewpoint, Ellis and Shintani (2014) considered error correction in response to feedback to be a form of uptake and an important step in the L2 acquisition process. They quoted Chandler (2003) who compared the increased accuracy of two groups of writers, one that received indirect feedback with no opportunity to revise and one that received indirect feedback with opportunity to revise. Chandler found that the accuracy of the group that had the opportunity to revise improved significantly more in a later piece of writing, concluding that uptake increases as a result of feedback with revision. Ellis and Shintani also quoted research by Van Beuningen, et al. (2012) who compared revision behaviour of a direct feedback group and an indirect feedback group. They found that there was increased accuracy in a second piece of writing for the direct feedback group only. From their own research into the use of the definite article, Ellis and Shintani found that lower-intermediate learners were not able to repair errors

based on feedback alone. They argued that the issue was not so much whether the feedback was direct or indirect but whether the writers were able to develop an understanding of the language rules connected with their errors.

From a sociocultural point of view, Lantolf and Poehner (2014) stated that when feedback is viewed from a purely interactionist perspective, research results are problematic because feedback on error needs to match the learner's ZPD and promote self-regulation. They argued that "[t]he question is not which form of feedback, implicit or explicit, is inherently better but which is most appropriate in the context of a particular interaction" (p. 172).

Polio (2012) argued that the above differences in research results arise out of the researchers' different theoretical beliefs. Interactionists view written corrective feedback as an essential tool for drawing the learner's attention to form; skill-building theorists see it as support during language practice; and socioculturalists consider it effective when it scaffolds learning within the ZPD and leads to self-regulation. Polio concluded that written corrective feedback on L2 writing, while time-consuming, is helpful to language learning when it is at the right developmental level for the learner, when the learner pays attention to it, and when the learner has opportunity to use the language immediately after correction.

Ellis and Shintani (2014) also surveyed corrective feedback research from different theoretical stances. Those who adhere to Universal-Grammar theories posit that feedback results in increased explicit knowledge, but it does not lead to the implicit knowledge associated with L2 acquisition. Those who adhere to cognitive-interactionist theories posit that feedback contributes to acquisition when learners can subsequently produce the corrected form with greater accuracy than previously shown. Those who adhere to sociocultural theory posit that feedback contributes to acquisition when learners are able to self-correct and use the form more and more consistently over time.

In the same way, Bitchener and Storch (2016) conducted an extensive survey of both interactionist and sociocultural feedback literature and concluded that: "there is not one type of written (corrective feedback) that can be identified as the most effective for all learners on all occasions" (p. 65). They insisted that, if feedback is to initiate L2 learning, it must vary according to "intervening individual internal (cognitive and motivation/affective) factors and individual external factors" (p. 33).

The above section has examined Manchón's (2011a) claim that the Noticing and Output Hypotheses and the Focus-on-Form and Skill-building Theories support her belief in the language learning potential of L2 writing. The claim was substantiated by empirical evidence for each hypothesis and theory. In association with Skill-building Theory, it also explored the important but controversial role of feedback when students are writing to learn language.

2.3: Assessing L2 Writing

Having reviewed literature on the scope of L2 writing research and on the language learning potential of L2 writing, this third section reviews literature on the assessment of L2 writing. It considers recent trends in L2 writing assessment, it revisits Dynamic Assessment, Assessment *for* Learning and portfolio assessment (see Sections 1.2.3 and 1.3.3), and it explores empirically-tested advice to teachers on the implementation of writing portfolios.

2.3.1: Recent Trends in L2 Writing Assessment

Yancey (1999) identified three waves of L2 assessment dating from 1950 to 2000: during the first wave, 1950-1970, writing was assessed through objective testing; during the second wave, 1970-1985, writing was assessed holistically in timed essays; and during the third wave, 1986 to 2000, writing was assessed by portfolio. Hamp-Lyons (2002), however, claimed four writing generations from the 1970s to 2000 and beyond. She reversed Yancey's first two waves. Her first generation was direct assessment and her second multiple-choice. She agreed with Yancey's third wave by stating that "[t]here is little disagreement that the last 15 years of the twentieth century turned the attention of writing assessment specialists and many other educators to portfolios as a fruitful form of assessment" (p. 10). In addition, Hamp-Lyons predicted a fourth generation of writing assessment for the first decade of the new millennium when online technologies would allow for the development of smart systems so a student writer would be able to build a writing portfolio and select elements from it according to need and purpose. She added her support to the principles of Assessment *for* Learning and praised teachers who make the effort to use "response to writing, peer feedback, self-reflection and self-assessment and all sound paths open to them to place writing assessment in support of teaching" (p. 13-14).

While Yancey (1999) and Hamp-Lyons (2002) chronicled twentieth century L2 writing according to a linear history, East (2008) preferred looking at L2 writing assessment from a theoretical viewpoint. East argued that the history of L2 writing assessment was not linear because L2

knowledge is often tested concurrently via external examination (large-scale timed tests) and internal assessment (coursework such as portfolios). East added that "[t]he two assessment paradigms are not mutually exclusive... Neither one is 'right' or 'wrong'. They are simply different, and based on different assumptions about what we want to find out" (p. 13). East suggested that timed tests have remained popular because of their reliability (the testing authority is able to control such variables as task, timing and marking methods), their validity (they test summative writing performance in a uniform and controlled manner and one performance can be evaluated against all other performances), their fairness (test-takers have the same opportunity to present evidence of proficiency), and their relative practicality, timeliness and cost-effectiveness. However, in communicative contexts where genuine social interaction is the goal, using learner-centred internal assessment to give learners the opportunity to convey real messages in authentic situations is also well established. East (2008) believed that the predominance of communicative methodologies has required language testers to move beyond testing language features in isolation towards testing language in terms of written proficiency.

In the New Zealand context, writing assessment has mainly followed the pattern that Yancey described, with multiple-choice style questions assessing reading and writing together until the 1980s when timed essays were introduced. Then, between 2002 and 2011, as indicated by East (2008), language proficiency was measured by both timed essays in external examinations and an internally assessed piece of process writing. After 2011, however, assessment by portfolio became the sole means of assessing writing and remains the current form of foreign language writing assessment today. This move towards assessment by process-based portfolio aligns well with the communicative methodologies encouraged by the MOE and with the principles of Assessment *for* Learning they have advocated since the 1980s[8].

2.3.2: Assessment *for* Learning, Portfolio Assessment, and Dynamic Assessment

According to Yancey (1999), Hamp-Lyons (2002) and East (2008), twenty-first century Assessment *for* Learning trends have resulted in a world-wide increase in outcomes-based and standards-referenced L2 curricula and a world-wide movement towards assessing L2 writing by

---

[8] It is interesting to note that Hamp-Lyons (2002) predictions that writing will be assessed dynamically and online are also beginning to come to pass. Since 2014, internally assessed writing portfolios are able to be stored and accessed for national moderation using various approved online technologies.

portfolio. Colby-Kelly and Turner (2007) described Assessment *for* Learning in the metaphor of an assessment bridge, "the place where assessment, teaching, and learning interweave in the classroom and … the place where language testing and second language acquisition interface, with the hoped-for result of learning" (p. 11-12). I. Lee (2007) listed some of the characteristics of learning programmes where Assessment *for* Learning principles guide teaching practice. They include sharing learning goals with students, helping students understand the standards they are working towards, involving students in assessment, providing feedback, creating a classroom culture where mistakes are a natural part of learning and improving, self and peer feedback, student-teacher conferences, and portfolio assessment.

In addition to the features listed by I. Lee (2007), several authors have emphasised the reflective elements of portfolios. Black and Wiliam (1998) described a portfolio as "a collection of a student's work, usually constructed by selection from a larger corpus and often presented with a reflective piece written by the student to justify the selection" (p. 45). They claimed that portfolios not only facilitate student-centred learning as students reflect on their own progress, but the process of designing reflection for student learning causes teachers to become more reflective themselves. In the introduction to her text advising teachers on the use of portfolios, Grift (2007) focused on the reflective nature of portfolio writing and on the role of metacognition in the writing process. She stated that it is only when students are challenged to reflect on their learning habits as they write that real progress can be made.

Belgrad (2013) gave several reasons for replacing written exit examinations with portfolio assessment in a National Writing Project conducted in the USA in 1983. She explained that the traditional examination (that is, Assessment *of* Learning) assumes that knowledge has universal meanings, that learning is passive and that it involves mastery of discrete and isolated bits of information. On the other hand, portfolios (which are examples of Assessment *for* Learning) view knowledge as complex and multifaceted, learning as active, and process as important as product. In addition, whereas high-stakes assessment has traditionally sorted and ranked students according to achievement, portfolio assessment "advocates focus on how the on-going processes of student inquiry capture the cognitive abilities that underscore successful achievement and engage students themselves as participants in the instructional design and authentic assessment of key learning events in subject areas" (p. 4). The teacher-researchers in her elementary school mathematics study wanted to know whether portfolios empowered their students to become more

engaged in the learning process. For seventeen weeks the students and their parents tracked

grades with portfolio evidence then completed a survey. The researchers found that portfolio

assessment promotes student engagement and increases parental awareness. To confirm these

findings, Belgrad (2013) suggested that "[s]tudent reflection on learning in portfolio processes

should become a focus of future research. Empirical studies are needed to test the belief that

engagement of students in reflective learning processes has the potential to *increase*

achievement, communication (voice), and motivation" (p. 12).

Akin to but different from Assessment *for* Learning is the sociocultural concept of Dynamic

Assessment (see also Section 1.3.3), which Feuerstein (Feuerstein & Feuerstein, 1999)

developed to assess and educate children who had survived the holocaust in post-WWII Israel. In

the face of so many culturally deprived children, he understood that psychological functions were

dynamic and "highly amenable to intervention and guidance" (Lantolf & Poehner, 2014, p.164).

Feuerstein reasoned that standard assessment events measure proficiency in a given task, but

what needs to be assessed is not task achievement but task transcendence. When a learner

transcends the immediate demands of a given task by transferring knowledge to a new task, then

learning has occurred. In summary, Dynamic Assessment is not a single event like traditional

assessment but a framework for day to day classroom interaction where a teacher tracks

performance across contexts; and it is an integration of assessment and teaching where teachers

"gain insights into learner abilities necessary to guide their efforts to move development forward"

(Lantolf & Poehner, 2014, p. 170-1).

Slavkov (2015) viewed the assessment portfolio as a tool which is able to meld both Assessment

for Learning practices as well as Feuerstein's Sociocultural Theory. He viewed the sequences of

planning, drafting, revising, editing and distribution as a form of continuous dialogic engagement

between students, peers and teachers. He stated that "[w]hile the process approach to writing

was not necessarily conceived as part of sociocultural theory, it is compatible with it, and can

indeed be regarded as one of its integral aspects" (p. 83).

2.3.3: Implementing Portfolio Programmes

Despite positive expectations by education administrators regarding the use of writing

portfolios, their implementation at classroom level is not always plain sailing. It must be asked to

what extent Assessment *for* Learning principles give way to Assessment *of* Learning principles during implementation.

Colby-Kelly and Turner (2007) researched the mandated implementation of a portfolio assessment programme in pre-university ESL classes in Canada. The researchers found that many participating teachers feared assessment was taking over teaching and that teachers were not using all the feedback options available to them. Instead of using feedback to benefit all students, they continued to give only discrete teacher-student feedback to individuals. They concluded that, despite ample explanation of the new curriculum, teachers were not viewing assessment as a part of the learning process and were not applying the principles as intended. Colby-Kelly and Turner therefore called for more professional development so that teachers might better understand the principles behind the approach and how to implement them.

I. Lee and Coniam (2013) also researched the implementation of Assessment *for* Learning programmes in EFL writing classrooms in Hong Kong's examination-driven secondary school system. Their research asked two questions of the 167 students and two teachers who volunteered for their study: how Assessment *for* Learning affected students' motivation and performance, and which factors facilitated or inhibited the teachers' implementation. I. Lee and Coniam found, in answer to the first question, that the Assessment *for* Learning programme brought about improved L2 writing but not necessarily improved motivation. Students indicated that this was due to the tension between the old and new ways of assessing. In answer to the second question, I. Lee and Coniam also found that teachers: successfully instructed students on planning to write; strengthened their pre-assessment instruction; shared learning goals and assessment criteria with students; used forms for feedback; and aligned assessment with instruction. However, due to time pressure, they were not able to allow students to write multiple drafts and engage in peer review on a regular basis. The researchers found that two factors facilitated successful implementation: the teachers' enthusiasm and willingness to work collaboratively, and their professional knowledge and skills. Equally, two factors impeded the implementation: societal expectations of time-consuming detailed feedback on error, and the awarding of summative scores. I. Lee and Coniam concluded that the teachers' understanding of the principles and practices of Assessment *for* Learning was critical to implementing successful Assessment *for* Learning programmes.

Just as Colby-Kelly and Turner (2007) found in Canada, and I. Lee and Coniam (2013) found in Hong Kong, New Zealand foreign language teachers have often voiced frustration that, while they might believe in the principles of Assessment *for* Learning, there are three main problems to their implementation in the day-to-day classroom. The first is time. As emphasised by Barnard and Campbell (2005), "by its nature, process writing is time-consuming" (p. 85). In the reality of the classroom, portfolio writing assessment consumes the time that in the past used to be given to the sort of formative writing practice that facilitated explicit language learning.

The second problem is the tension between optimal learning conditions and reliable and valid assessment. East (2016) mused that while the authorities seem to encourage Assessment *for* Learning in principle, they "in practice encourage a testing model, not necessarily deliberately, but consequentially by virtue of the conditions surrounding internal assessment" (p. 197). He added that "collecting 'real life' evidence that emerges from students' regular work challenges fundamental notions of standardisation and reliability that traditionally inform high-stakes assessment" (p. 201). For example, Bitchener (2008) suggested that the most effective forms of feedback for language learning are explicit, written and immediate. However, in order for feedback to be fair in the context of assessment, the 2015 NZQA Clarifications specified that feedback should be indirect, and, in 2016, they specified that it should be holistic and given only once (NZQA, 2015; NZQA, 2016).

The third problem is the crucial role of the teacher. It must be conceded that, just as Colby-Kelly and Turner found in Canada, and I. Lee and Coniam found in Hong Kong, perhaps New Zealand teachers, for many optional or imposed reasons, are not taking full advantage of the sociocultural features of portfolios that promote learning, such as collaborative review, feedback that benefits many students rather than one, scaffolded and repeated practice, and the opportunity to reflect (Barnard & Campbell, 2005; East, 2016). When such features are omitted from process writing programmes, the language learning potential of writing is arguably diminished.

Because of the wide acceptance in principle that writing is best taught as a process rather than as a product (Barnard & Campbell, 2005), it is important to find ways to solve the time, tension and teacher issues detailed in the previous paragraphs. From a sociocultural perspective, Barnard and Campbell suggested that, where classes are large, two ways to deal with time issues when process writing is for students to collaborate on their writing in groups and to harness the benefits

of online conferencing technologies. They added that, like their students, teachers also need scaffolding in the effective implementation of process writing.

Based on her involvement with the implementation of Assessment *for* Learning in Hong Kong secondary schools, I. Lee (2007) also acknowledged that process writing with multiple drafts is time-consuming so she recommended that some pieces of writing should be single draft and not count for assessment. She also suggested empowering and motivating students with self- and peer-evaluation, and with error logs and reflective journals that detail how to improve future compositions. I. Lee found that, when students participate in the teaching and learning portfolio process by compiling their own portfolios, developing feedback forms or checklists and selecting specific features for teacher feedback, they are more likely to feel empowered and remain motivated.

East (2016) researched the implementation of a spoken interaction portfolio for high stakes assessment in New Zealand secondary schools. Over a two year period he surveyed 152 teachers and 149 students, and interviewed 27 teachers. As a result of the data gathered, he found conflicting stake holder perspectives that needed resolution. For example, the data revealed conflict between what is optimal for learning and what constitutes valid assessment. East argued that, when portfolios are seen from a sociocultural perspective, even in an assessment context, they are able to provide the sort of scaffolding that leads to L2 independence and automaticity. He suggested that this might include students working collaboratively to prepare an assessment task, and students being offered the opportunity to practise and repeat a task in response to feedback. While focused on portfolio assessment of L2 spoken interaction skills, East made several recommendations to teachers that might apply equally to portfolio assessment of L2 writing skills: remember that the ultimate goal of language learning is real and automatic communication with L1 speakers; offer as many opportunities to develop the skill as possible; provide feedback to enhance performance; assess performance in real world tasks; assess language according to what is appropriate for the text-type; foster a sense of ownership in students by offering more choice; and remember that natural language is not necessarily perfect language.

After researching the implementation of portfolio assessment in the USA with fifteen FFL university students, Paesani (2006) also offered advice to teachers. She suggested that teachers

emphasise with their students the three distinct processes involved in writing for portfolio: pre-writing, writing, and post-writing. Where pre-writing prepares students for the first draft and can involve brainstorming ideas, compiling vocabulary lists or developing an outline, and writing involves converting ideas to language and completing a draft with purpose and audience in mind, post-writing involves revising and rewriting until a final draft emerges. She underscored the importance of feedback from peer, self, or instructor for encouraging "collaboration, negotiation of meaning, critical thinking and an attention to multiple audiences" (p. 621).

Padilla, Aninao and Sung (1996) researched the implementation of L2 writing portfolios with approximately one thousand students learning Japanese, Chinese, Korean, and Russian in Californian elementary and high schools as well as universities. As a result of their research, they recommended writing portfolios as particularly advantageous in L2 contexts because, compared with portfolios in other subject areas, they revealed progress over time. The researchers reminded L2 teachers that the purpose and audience of the portfolio should decide what goes into it, who puts it there and how often, as well as who is responsible for its safe-keeping. They suggested that portfolios should contain a table of contents where the date, skill, topic and type of activity for each contribution are recorded as well as a chronological number so that progress over time is more clearly seen and drafts are not confused. Padilla, Aninao and Sung also suggested that older students might record reflections that explain why they chose a given artefact and detail the learning that the artefact initiated.

The above section has surveyed literature on the trends in L2 writing assessment, finding that the adoption of Assessment *for* Learning principles has led to an increase in the use of portfolio assessment in education systems around the world. Theory and advice from portfolio researchers in Canada, the USA, Hong Kong and New Zealand suggested that three key components of portfolio assessment are process writing, conversations around feedback, and reflective learning practices. The next section looks specifically at literature on reflective process writing practices in L2 contexts to establish why teachers might be encouraged to include them in their L2 writing programmes.

2.4: Reflecting when L2 Writing

This section explores what research literature says about reflection on L2 writing. It begins with the premise that process writing and reflection are co-dependent activities. Researching in an

L1 context, Scardamalia, Bereiter and Steinbach (1984) described the relationship in this way: "Most modern approaches to composition instruction give an important place to reflective processes, in contrast to the linear procedures often espoused in older composition textbooks. Not only is reflection valued as an aid to writing, but writing is valued as an aid to reflection" (p. 173).

In the ideal classroom described by the Assessment Reform Group (I. Lee, 2007), teachers create learning environments where teaching, learning and assessment are integrated, using such tools as the writing portfolio. As L2 learners write texts for their portfolios, teachers are able to provide students with feedback and time to reflect on that feedback. In the L2 classroom, writing provides the ideal occasion for reflective practices because students have the time to consider their language production and to strategise to improve, which is not generally the case when they engage in L2 speaking. In order to discover how reflection aids L2 process writing, the next four sections survey literature from the fields of Metacognition, Language Learner Strategies, and Computer-Assisted Language Learning.

2.4.1: Metacognition and Self-regulated Learning

Writing from the field of psychology, Flavell (1979) coined the term 'metacognition' to refer to a learner's awareness of learning processes. He theorised that there are three types of metacognitive knowledge, person knowledge (beliefs about universal truths and self-concept), task knowledge (knowledge about the general process and nature of learning), and strategic knowledge (perceptions about the use and usefulness of certain strategies).

Twenty years later, writing from the field of SLA, Wendon (1999) theorised that metacognition is composed of metacognitive knowledge (what learners know about their language learning) and metacognitive strategies (the skills that help learners manage and regulate their language learning, such as planning, monitoring and evaluating).

A decade further on, Houston and Turner (2007) theorised about metacognition from a Mindful Learning approach. They claimed that the concept of mindfulness has much to offer SLA theory, and vice versa. They described mindfulness as the ability to determine how to use knowledge and skills and the ability to choose which outcomes are meaningful and why. Consequently, a mindfulness approach enables L2 learners to step back and exert control over their learning and apply meaning to their learning outcomes. As learners mindfully manage their L2 input through

interaction, their awareness of form is heightened, which pushes them to process the input actively. A mindful approach to language learning also emphasises outcomes that are meaningful to the learner, so any grammar instruction will be meaning-based and will emphasise a meaningful outcome.

Houston and Turner's (2007) work was in response to the work of Langer (2000) who wrote from the field of psychology. Langer described mindfulness as "a flexible state of mind in which we are actively engaged in the present, noticing new things and sensitive to context ... (and) when we are mindful, rules and routines may guide our behaviour rather than predetermine it" (p. 220). Through her research, she found that when participants were asked to pay attention to a phenomenon and to notice new things about that phenomenon, in other words, to be mindful about it, they remembered more about the phenomenon and they liked the phenomenon better. She concluded that "[m]indful learning engages people in what they are learning, and the experience tends to be positive" (p. 222).

Knouzi, et al. (2010) conceptualised metacognition in the context of L2 writing as self-scaffolding and self-explaining or, following Swain (2006), as 'languaging'. As also described in Section 2.1.2, Knouzi and her colleagues conducted a microgenetic analysis of the languaging behaviour of two young adults studying FFL to find out why languaging helps some learners more than others. They defined languaging as students using language "to intentionally organise and control their mental processes during the performance of cognitively complex tasks" (p. 24). Self-scaffolding in the form of languaging might involve organising one's own learning, breaking down problems into manageable chunks, and choosing resources to help understanding. The researchers found that, through speaking to themselves, students made linguistic discoveries and connections that they were not previously aware of; and they mediated their own learning processes which resulted in increased language learning. In addition, when students used languaging to solve linguistic issues, there was a movement of scaffolding agency away from the teacher towards the learner. Knouzi, et al. therefore recommended that teachers create "time and space for individual students to explore and develop self-scaffolding tools during and outside instructional time" (p. 46). Teachers might do this by waiting longer for students to think before answering questions, talking less and encouraging students to talk to each other more, and modelling languaging as a useful tool for L2 development.

Chamot, et al. (2005) defined metacognition as "awareness of one's own thinking processes" (p- 149) and argued that this self-knowledge is valuable because it leads to independent learning:

> (Metacognition) leads to reflection, to planning how to proceed with a learning task, to monitoring one's own performance on an on-going basis, and to self-evaluation upon task completion. In other words, it leads to self-regulation of one's learning. Students with greater metacognitive awareness understand the similarity between the current learning task and previous ones, know the strategies required for successful learning, and anticipate success as a result of knowing 'how to learn' (p. 149).

From the field of educational psychology, Zimmerman and various colleagues over four decades have conducted multiple studies that have resulted in Zimmerman's theory of six dimensions of self-regulated learning (Schunk & Usher, 2013; Zimmerman, 1990). As illustrated in Table 2, in the dimension of self-regulation called Time, Zimmerman emphasised the importance of time management to learning; in the dimension of self-regulation called Method, he emphasised the role of strategies and routinised performance in learning; in the dimension called Motive, Zimmerman emphasised the importance of setting goals and the self-efficacy that goal setting produces; in the dimension called Behaviour, he emphasised the importance of self-observation, self-judgment and self-reaction for learning, and their key strategies of monitoring and keeping records or self-recording; and in the dimensions called Physical Environment and Social Environment, Zimmerman emphasised the ability of self-regulated learners to structure their learning environment, and to network and seek help from others (Schunk & Usher, 2013; Zimmerman, 1990).

Table 2: Zimmerman's Self-Regulated Learning Dimensions (Schunk & Usher, 2013, p. 9)

Dimension	Key Processes
Time	Time management
Method	Strategies; Routinised performance
Motive	Goals; Self-efficacy
Behaviour	Self-observation; Self-judgment; Self-reaction
Physical Environment	Environmental structuring
Social Environment	Social Networking; Selective help seeking

A first step in the self-regulation process is self-monitoring, and a simple and important way to self-monitor is to track learning by recording it. Noting goals, strategies, and response to feedback, and tracking progress regularly and immediately after participating in an activity is

especially effective because it provides an opportunity to enhance motivation and self-efficacy and thereby promote continued learning (Graham & Macaro, 2007; Schunk & Usher, 2013).

In their research to determine the use of self-regulated learning strategies by high school students, Zimmerman and Martinez-Pons (1986) found that higher track students use more strategies than lower track students and the four strategies that distinguish the best students include keeping records and monitoring progress. Williams and Burden (1997) suggested that, when language teachers provide students with tools, such as questionnaires and journals, to record and evaluate their learning, they are applying Feuerstein's theory of mediation; in particular, they are ensuring the intentionality-reciprocity that is essential for learning to occur. When they apply these mediations in a systematic manner, "their influence can be both powerful and profound" (p. 84).

Whether from the field of psychology (Flavell, 1979), from a mindful approach (Houston & Turner, 2007; Langer, 2000), from cognitive-interactionist perspectives of SLA (Wendon, 1999), from sociocultural perspectives of SLA (Chamot, et al., 2005; Knouzi, et al., 2010), or from the field of self-regulated learning (Schunk & Usher, 2013; Zimmerman, 1990), scholars agree on the significant links between metacognition and learning. It is therefore important that L2 educators make the time and the occasion for students to reflect as they write so that the language learning potential of writing is increased. However, making time and occasion is not always enough; secondary school students, in particular, also need guidance and support. The next sections consider what guidance on reflection might look like in terms of topic, language and mode.

2.4.2: Possible Topics for Reflection

The literature already surveyed has raised several topics for reflection when L2 writing: learning habits (Belgrad, 2013; Grift, 2007), goal setting (Feuerstein & Feuerstein, 1999; I. Lee, 2007), metalinguistic reflection on feedback (Bitchener, 2008; Manchón, 2011a), and learning and communication strategies (Anderson, 2002; Chamot, et al., 2005; Oxford, 2011; Zhang, 2013). Each is now revisited in turn.

The first possible topic for reflection is learning habits. Grift (2007) maintained that it is only when students are challenged to reflect on their learning habits that real progress in learning can be made. Belgrad (2013) called for empirical studies to test the belief that when students engage in reflective learning habits their achievement and motivation increases. Claxton (2006) divided

learning habits into mentality habits, the "cognitive habits of mind that go to make up one's capacity to learn," and identity habits, "the emotional and personal attitudes, beliefs and tolerances that expand or contract one's capacity to learn" (p. 5). He argued that schools have a responsibility to create a climate in which students feel continually empowered, not ignored, and where students know *how* to learn as well as *what* to learn. To achieve this, he suggested regular plenary and small group discussions as well as "reflective writing in learning diaries" (p. 9).

A second possible topic for reflection is goal setting which is listed by I. Lee (2007) as one of the characteristics of Assessment *for* Learning. Goal setting is also a feature of Feuerstein and Feuerstein's (1999) sociocultural model of mediated learning. It links to the essential mediation of transcendence. Teachers promote transcendence when they ensure their students understand the long-term, transferable purposes of their immediate learning activities; they can do that by encouraging their students to set and achieve learning goals (see Sections 1.3.3 and 2.3.2).

Oxford (2011) categorised goals as either mastery (those related to developing competence) or performance (those related to demonstrating competence compared with others). Her research indicated that, in situations of certainty, performance goals lead to improved performance but, in situations where there is uncertainty, such as assessment, performance goals are not motivating at all. Her research also indicated that students who set mastery goals have stronger motivation, choose more cognitively challenging goals, and persist longer. Because both mastery and performance goals can be positive under positive circumstances, and are more effective combined than either is alone, Oxford argued for a multiple goals approach. Language students who focus on both the short-term performance goal of examination success as well as the long-term mastery goal of fluency are therefore more likely to persevere, continue learning, and succeed in achieving both goals.

The links between goal setting, self-efficacy, motivation and achievement are clearly seen in Schunk and Zimmerman's (2009) definition of self-regulation: "the control of one's present conduct based on motives related to a subsequent goal or ideal that an individual has set for him- or herself" (p. 1). Schunk and Zimmerman found from their research that good self-regulators are better at setting learning goals, implementing learning strategies, monitoring and evaluating their progress, maintaining effort, and, having achieved their goals, at setting new goals. They also found that students who set learning (mastery) goals "recovered more quickly from poor

performance on the first exam in the course and displayed higher performance by the end of the course than students with a performance goal orientation" (p. 7). They concluded that to set learning goals is to believe in learning potential rather than fixed intelligence, a belief that motivates non-confident as well as confident learners alike.

Gordon (2008) maintained that becoming a proficient L2 writer requires consistent and persistent effort over a long period of time. An important key to maintaining the required effort is goal setting, whether extrinsic or intrinsic. In a five-year intervention study that took place in 23 secondary schools in Nebraska with 1,273 Spanish as Foreign Language students, Moeller, et al. (2011) researched the connection between goal-setting, reflection and achievement levels through the implementation of LinguaFolio, "a standards-based, self-directed, formative assessment tool designed to increase learner autonomy through a carefully structured goal-setting process" (p. 153). They found a statistically significant correlation between the students' goal setting and reflection process and language achievement levels.

To increase self-efficacy, motivation and achievement, therefore, L2 educators can support their students in goal setting by giving them the time and the guidance they need to set and reflect upon their learning goals. Williams and Burden (1997) suggested a simple form with sentence starters such as: "By the end of this course/term I want to … My plan to achieve this is ..." or "This week I shall …" (p. 76). In order to facilitate students' awareness of moving towards their goals, they suggested that learners monitor their progress with logs of information learned, skills perfected, or best work. Another option is a form where students record learning activities and the reasons why they did them. Williams and Burden claimed "[a]n important consequence of learning how to monitor one's own progress in learning is that one comes to recognise personal change as continuous, lifelong and within one's power" (p. 76). Goal-seeking, setting and achieving is therefore a likely contender for student reflection when writing.

After learning habits and learning goals, a third possible topic for reflection in an L2 writing context is metalinguistic reflection on feedback. As indicated in Section 2.2.3, researchers have often disagreed about the role of feedback in language learning. For example, Oxford (2011) found teacher feedback to be often confusing and inconsistent and therefore ignored by students; Ferris (2006) found teacher feedback on error to be incorrect in only 3.6% of instances; and Hyland (2011) viewed teacher feedback as potentially one of the most influential aspects of process

writing as the teacher intervenes to pay individual attention and to offer overt correction and explicit teaching.

Traditionally, feedback has been given by the teacher to the student but, in a portfolio writing context, it might also be peer-feedback or auto-feedback. From a sociocultural point of view, it would seem that teacher feedback on writing is more appropriate at the intermental stage of microgenetic development (Aljaafreh & Lantolf, 1994) and that auto-feedback is more likely to occur at the intramental stage of development (Swain & Lapkin, 1995), with peer feedback being more appropriate in the middle stages of microgenetic development. Knouzi, et al. (2010) found that, as learners come across problems in their writing and engage in languaging to overcome those problems, whether with a teacher, a peer or themselves, this prompts the reflective processes that are capable of building new knowledge.

Manchón (2011a) distinguished between feedback for accuracy and feedback for acquisition. She defined feedback for acquisition as the sort of feedback that prompts deep linguistic processing. This is most likely to occur in writing programmes where learners receive differentiated feedback specific to their ZPD, and are given time to reflect on this feedback, to notice the gap between their own knowledge and this new knowledge, and then to produce this new knowledge in subsequent drafts. Kuiken and Vedder (2008) concluded from their research that the depth of the metalinguistic processing that occurs while students write, whether simple or elaborate, will determine the significance of their L2 outcomes.

In addition to reflection on learning habits, goals and feedback, a fourth possible topic for reflection is language learner strategies. Gardner (1963), a pioneer in self-regulated learning, insisted that the goal of education was to "shift to the individual the burden of pursuing his own education" (p. 21). By observing learner behaviour, three decades later, Zimmerman (1990) defined three behaviours that set self-regulated learners apart: their use of self-regulation strategies, their responsiveness to feedback, and their ability to motivate themselves. Over several studies, Zimmerman was able to compile a list of the fourteen learning strategies that distinguished self-regulated students: "self-evaluation, organisation and transformation, goal setting and planning, information seeking, record keeping, self-monitoring, environmental structuring, giving self-consequences, rehearsing and memorising, seeking social assistance (from peers, teachers, or other adults), and reviewing (notes, books, or tests)" (p. 7).

In the field of SLA, Gu (2010) argued that "strategic, self-regulated learning lies at the heart of Second/Foreign Language acquisition" (p. 10). He maintained that linguists have been preoccupied for decades with such issues as comprehensible input, output, corrective feedback, task-based language learning and scaffolding. Now researchers are turning their attention to the language learners themselves and the role they can play in controlling their own learning and maximising their learning outcomes with support from their teachers.

Anderson (2002) agreed:

> The teaching of metacognitive skills is a valuable use of instructional time for a second language teacher. When learners reflect upon their learning strategies, they become better prepared to make conscious decisions about what they can do to improve their learning. (p. 1)

Anderson (2008, p. 99) divided metacognition in language learning contexts into five primary but intersecting elements that are not dissimilar to the processes of portfolio writing: 1. preparing and planning for learning; 2. selecting and using strategies; 3. monitoring learning; 4 orchestrating strategies; 5. evaluating learning.

After forty years of collaboration and research, Oxford (2011) compiled an extensive framework of language learner strategies which she divided into three dimensions: cognitive, affective, and sociocultural-interactive. She listed strategies in the sociocultural-interactive dimension as those that deal with issues of contexts, communication, and culture in learning, such as, interacting to learn and communicate, overcoming knowledge gaps in communicating, and dealing with sociocultural contexts and identities. The affective strategies she listed as those that handle emotions, beliefs, attitudes and motivation, such as, activating supportive emotions, beliefs and attitudes and generating and maintaining motivation. The strategies she listed in the cognitive dimension are those that help students remember and process the L2, such as, using the senses to understand and remember, activating prior knowledge, reasoning, conceptualising with details, conceptualising broadly and going beyond the immediate data. She listed meta-strategies as those strategies that allow students to manage their L2 learning by understanding their own needs and using and adjusting the other strategies to meet those needs, such as, paying attention, planning, obtaining and using resources, organising, implementing plans, orchestrating strategy use, monitoring, and evaluating.

When planning for strategy instruction, Cohen (2011) reported that age and proficiency are important variables to be considered. Whereas beginners tend to favour less effective strategies (e.g. word by word translation) more experienced learners tend to use more effective strategies. However, as language becomes more fluent, fewer strategies are used because language learning has become more automatic and less strategic. Cohen and Macaro (2007) noted that, although ineffective learners tend to lack the range and combinations of strategies of more effective learners, they particularly lack the metacognitive strategies. Ramesh (2009) agreed that understanding and controlling cognitive processes may be one of the most valuable skills a classroom teacher can develop in L2 learners because using metacognitive strategies can lead to deeper learning and improved performance. Setting time aside for these learners to reflect in a way that they might not otherwise do has the potential to increase their capacity for learning.

In an analysis of British students and their language learning choices, Graham (2004) found that those with higher levels of achievement, and those who intended continuing to learn a language, were most likely to attribute their success to their effort, their higher ability, and the learning strategies they used, whereas those who struggled to achieve were more likely to cite lower ability and task difficulty for their lack of achievement and discontinued learning. They failed to mention learning strategies. Graham argued that the learners' confidence in their potential to succeed and their desire to continue learning might be improved if they saw the links between learning strategies and academic performance.

Following up Graham's (2004) research with a nation-wide intervention study, Graham and Macaro (2007) created a strategy instruction programme for which they compiled formulation strategies suitable for use by their Year 12 FFL participants. Because their research suggested that planning was almost completely missing from the students' repertoire, the first strategy taught was brainstorming or mind-mapping. The researchers wanted to encourage students to retrieve from long-term memory phrases that matched the task requirements. This required thinking in French. The second strategy taught was evaluating the gap. After brainstorming in French, students were taught to consider how much language they still needed to respond to the task requirements. As they began to write, the students were taught a set of strategies to overcome the gaps in their knowledge, for example, combining one set phrase with another, restructuring a set phrase by changing parts of it, translating from English to French, consulting a bilingual dictionary, or even avoiding the sentence altogether. A further set of strategies which related to

monitoring or checking their writing included visual monitoring, auditory monitoring, backtranslating, or prompting with a reminder, a rhyme or a mnemonic. The last set of strategies taught related to responding to feedback. To encourage the notion that writing is about communicating meaning, they asked students to comment on each other's content, gave them practice exercises related to their errors, and offered feedback on the students' strategy use as well as their accuracy. Due to the secondary school students' level of French, this strategy training took place in English.

When contemplating possible topics for reflection that are likely to promote L2 learning, learning habits, goals, metalinguistic processing of teacher feedback and language learner strategies are all likely contenders. Across the course of a secondary school year, or even within one writing unit, it seems possible that students might be guided, at appropriate moments, to reflect on all of these.

2.4.3: Language of Reflection

Having considered possible topics for the sort of reflection that has the potential to promote language learning, a second important consideration is the language of reflection - should it be the L1 or the L2? Because of the continued popularity of communicative methodologies, strong beliefs are held regarding the use of the L1 in the L2 classroom (Turnbull & Daily-O'Cain, 2009). These beliefs may have developed as a result of the successful Canadian immersion programmes where a core principle is exclusive target language use. In some contexts, the L1 has even been officially banned in the desire to avoid code-switching (changing from one language to another within the same sentence or conversation) and to minimise L1 interference. Proponents of L2-only justify their stance by quoting the Output Hypothesis which recognises that, although comprehensible input is important, in order for learners to master the target language, they need ample opportunity to produce it (Turnbull & Daily-O'Cain, 2009). Therefore, if language learning is the goal, should not reflection be in the L2?

In their research, Cohen and Macaro (2007) found that small amounts of strategic L1 use in the classroom contributed to more comprehensible input and more target language production, whereas L2-only insistence resulted in too much teacher-talk and less L2 production. Cohen and Macaro considered this issue to be the most interesting feature of L2 writing research in recent years. They found that, in the beginning stages of L2 writing, the L1 is used as a compensatory

strategy: learners think, access ideas, access words, and evaluate those words with backtranslation to judge their appropriateness, all in the L1. As the learner gains more control over the L2, the L1 is still used when problems arise. It is used to frame those problems and access and monitor meanings and forms stored in memory. As fluency in writing increases and the L2 becomes more automatic, some higher level processing, such as, planning, organising and solving rhetorical and discourse problems still occurs in the L1. Even at the highest levels of L2 proficiency, the L1 might still be used for task conceptualisation, planning, organisation, and monitoring and evaluating the writing process.

When conducting research with four L2 French students who were interacting in online blogs, Lamy and Hassan (2003) noted their use of the L1 and its link to reflection and deep processing. They concluded that, when the L1 was used for linguistic comparisons between the L1 and the L2, online discussion led to talk of a more reflective nature, and when participants helped each other with technology problems, negotiations over terminology were triggered and long-term vocabulary retention was the result. This confirmed their hypothesis that there is a strong link between spontaneous interaction, L1 and L2 comparisons, and deep learning, especially when they involve problem-solving and expertise sharing.

From a sociocultural perspective, language is a cognitive tool which mediates all forms of higher-order thinking. In contexts where the students' level of L2 knowledge is low and where the cognitive load of the task is heavy, there is considerable benefit in allowing L1 use to reduce working memory constraints (Turnbull & Dailey-O'Cain, 2009). Ellis and Shintani (2014) maintained that learners will most often use their L1 when involved in private speech because the audience is themselves. To use the L2 for private speech is difficult for even the most advanced of learners. Ellis and Shintani concluded that while it is clear from research that effective language learning requires extensive L2 input and opportunities for output, there is a strategic role for L1 use in the L2 classroom. Nevertheless, they stated that "there is a conspicuous lack of research that has investigated what effect (facilitative or debilitative) use of the L1 has on actual learning" (p. 245). They recommended that teachers reflect carefully and critically on their L1 and L2 use in the classroom.

Swain and Lapkin (2000) did, in fact, research the L1 use of 22 pairs of Grade 8 French students in a Canadian immersion school as they completed two different writing tasks. They were

surprised to find that a quarter of the interactions took place in the L1. On closer examination they found that only 12% of the L1 utterances were off-task communication, the rest served important cognitive and social functions for L2 learning, such as understanding the requirements of the task and task organisation, paying attention to language form and vocabulary use, and establishing a collaborative working relationship. Swain and Lapkin (2000) concluded that "judicious use of the L1 can indeed support L2 learning and use. To insist that no use be made of the L1 in carrying out tasks that are both linguistically and cognitively complex is to deny the use of an important cognitive tool" (p. 268-9).

Based on her research, Laoire (2014) also drew conclusions regarding the use of L1 English. She used L1 interviews to investigate the extent to which L2 learners of Irish in five Irish secondary schools were able to transfer skills, strategies and experiences from their compulsory L2 Irish learning to their optional L3 French, German or Spanish learning. Laoire found that the learners' metalinguistic knowledge was superficial and lacking in reflection and engagement. She attributed this to the unhelpful tradition of teaching Irish via a direct method that did not allow any use of English at all.

Manchón, et al. (2009) stated that "switching to the L1 is without doubt one of the most characteristic features of L2 writing" (p. 114). They found that L2 writers use their L1 for many strategic purposes related to understanding the task, planning, formulating and reviewing, and monitoring the writing process. Interestingly, because the L1 is used more in the planning and reviewing stages of writing, and because less proficient learners spend more time formulating than planning and reviewing, they found that more proficient learners use their L1 more. They concluded that "[f]oreign language writing is certainly a multilingual event" (p. 121).

2.4.4: Modes of Reflection

Having explored literature on topics for reflection and the appropriate language of reflection, it is important to now ponder the modes of reflection. Warschauer and Healey (1998) suggested that the natural choice of mode for twenty-first century learning is technology. Like Hamp-Lyons (2002), they predicted that the first decade of the new millennium would feature intelligent Computer-Assisted Language Learning and that there would be an increased emphasis on electronic literacies, online writing and artificial intelligence in L2 learning contexts. In such an environment, learners would have access to interactive, intelligent multimedia that gave corrective

feedback, that fitted the learner's learning styles, and that facilitated interaction with the world. They warned that second language teachers would need to teach the genres of electronic communication and the relationship of text to other media without diluting attention to language. Warschauer and Healey indicated that "[a]s our focus of attention gradually shifts from the computer itself to the natural integration of computers into the classroom learning process, we will know that computer technology has taken its rightful place as an important element of language learning and teaching" (1998, p. 67-68).

The Bring Your Own Device (BYOD) policies that many secondary schools in New Zealand have implemented in recent years (62% in 2015 according to the NZCER) have assured that (in a sociocultural sense) technology is not only integrated but also appropriated by schools (Williams & Burden, 1997). When learners access textbooks and exercise books online, it seems logical to assume that they are going to reflect online as well. When learners conduct all aspects of their learning digitally, to consider the impact of reflecting using technology becomes redundant. However, many advantages to reflecting online can be noted: the time-saving convenience of digital templates, word processing and editing, the security of digital storage, copying and sharing, the ubiquity of 'anyone, anytime, anywhere' access, the convenience of synchronicity and asynchronicity, and the motivation and variety that multimedia provide.

Based on their research, Lamy and Hassan (2003) offered advice to teachers when designing online activities to elicit the sort of reflection that results in language learning. They suggested "building into any design 'a psychological and conversational space' in which learners can be responsible for task-management, as 'themselves'" (p. 54). They found that their four research participants were more reflective about language and language learning when they had the space to discuss their own study arrangements, their past experiences with learning, and their current learning beliefs. Lamy and Hassan also emphasised the importance of explicit instructions when students reflect online. This is especially important with secondary school language learners who do not necessarily take the initiative to reflect because they prefer to look forward rather than backwards (Porto, 2007).

When researching computer-mediated-discussion through the Open University, Lamy and Goodfellow (1999) found that asynchronous online technologies were better for encouraging metalinguistic reflection than synchronous technologies because they give learners the time they

need for thinking. Based on their study, they found that using asynchronous technologies to facilitate reflective conversation on language learning is particularly beneficial for negotiation of meaning, attention to form, and strategy use.

Hyland (2011) suggested that structured diaries give students guidelines to help them produce targeted data and at the same time provide them with the metalanguage they need to talk about their writing processes and strategies. Adolescent learners especially require the scaffolding that structured diaries provide. However, in order to prevent diary keeping from becoming a burden, researchers must devise strategies to motivate participation (Hyland, 2011). The inclusion of habit forming times set aside for reflection as an integral part of writing lessons is one such strategy. Another is the provision of a customisable, digital template and online space for saving reflections alongside writing samples.

However, Griffiths (2008), in her study of the language learning strategies used by 131 young international ESL students in a New Zealand study skills programme, found that diary writing was used infrequently as a strategy for learning and it was infrequently used by both higher and lower students alike. She therefore suggested that "teachers might like to consider the implications of this finding for their own practice and could very usefully engage in some action research to investigate whether students who write diaries progress more quickly than those who do not" (p. 88).

Rubin (2003) suggested using diaries to teach students how to develop a metacognitive awareness of their learning processes and strategies. She viewed the use of diaries in foreign language classes as particularly powerful tools because they motivate students and allow them to "begin to take control of their learning" (p. 14). She continued:

> The more experience learners have with diary writing, the better they become at it and the more reflective they become. As research tells us, expert learners are in charge of their learning and know how to change the process to suit their learning styles, language and background knowledge, and the task. (p. 14)

Anderson (2008) agreed that language learning journals are excellent tools to give to students to encourage reflection. He added that, when shared, they also allow teachers to gain insights into their students' metacognition. He advocated that teachers provide prompts to reflection in learning

journals so that students are supported and encouraged to engage in all aspects of metacognition that enhance language learning in the context.

Murray, Hourigan and Jeanneau (2007) conducted a four year study on academic writing using blogs as language learning journals. One activity involved students sharing their language learning strategies and experiences. The goal was improved language learning through self-reflection and self-expression. The researchers found through a survey that most students understood the language learning potential of diary-like blogs when used as reflexive tools.

Martinez (2012) researched undergraduate Spanish students as they used blog technologies to reflect on language learner strategies. The activity was designed to promote metacognition and learner autonomy. Martinez found the blog to be a suitable complement to traditional face-to-face courses and a successful vehicle for students to share strategies and experiences that promote L2 learning.

Slavkov (2015) married the rationale of sociocultural theory and process writing with the use of Google Drive. Because process writing is socially constructed and dialogic in nature, he claimed that it is a perfect fit with sociocultural theory and also with Google Drive as a twenty-first century technology. Slavkov argued that this platform is capable of playing a significant supporting role in sociocultural process writing because it offers "document creation, storage, synchronous and asynchronous sharing, editing and commenting, revision histories, document-specific real-time chat, and various distribution operations across audiences and platforms" (p. 83). It also allows writers to select their sharing audience which makes it an ideal online platform for reflection.

2.5: Summarising the Chapter

This chapter surveyed research literature to establish what has already been said about L2 writing in order to discover directions for further research. The review included sections on the scope of previous L2 writing research, the language learning potential of L2 writing, the assessment of L2 writing, and reflection on L2 writing.

In the first section, Manchón's (2011a) categories of L2 writing research guided a review of sixteen studies relevant to the present study. Torras and Celaya (2001) used intergroup and intragroup analysis, Ruiz-Funes (2015) and Torras and Celaya (2001) used a range of CAF measures, Roca de Larios, et al. (2008), Ruiz-Funes (2015), and Sasaki (2000) divided

participants into three achievement levels for finer measurement, and Adams (2003) used a seven day timeline between feedback and rewriting. In process and strategy research, Ong (2014) established the importance to ideas and language of giving students planning time when L2 writing, and De Silva and Graham (2015) established the positive impact of strategy training on both high and low achievers and discovered the effectiveness of immediate stimulated recall to gather data. In feedback studies, Aljaafreh and Lantolf (1994) and De Guerrero and Villamil (2000) measured microgenetic development, Bitchener (2008) used one pre-test and two post-tests in his research design, and Lee (2015) found that, while junior secondary school students prefer teacher feedback, they enjoy and gain confidence from peer feedback. Izumi (2002) reiterated that learning when writing occurs through noticing and producing noticed forms, Knouzi, et al. (2010), Lamy and Hassan (2003), and Laoire (2014) established the link between reflective practices and depth of linguistic thinking, Laoire (2014) saw the importance of using the L1 to reflect on language acquisition, and Teng and Zhang (2016) established the connections between self-regulated learning, writing strategies and written performance.

The second section surveyed L2 writing through studies that illustrated the Noticing and Output Hypotheses, Focus-on-Form, Skill-building Theory, and corrective error feedback. These studies confirmed Manchón's (2011a) claim that writing has the potential to promote language learning when learners are given the time and the guidance to notice linguistic forms and to reflect in such a way that deep, explicit, metalinguistic processing occurs. When students read, plan, compose, reflect and edit in response to feedback, they construct an environment in which L2 learning becomes possible (Manchón, 2011a). Despite longstanding controversies regarding the effectiveness of feedback, most recent research concludes that feedback promotes language learning when it differentiates for individual learning needs, when opportunities for rewriting immediately follow, and when "learners make use of the corrections they receive to develop metalinguistic understanding of the nature of the error and why it is corrected" (Ellis & Shintani, 2014, p. 274).

The third section surveyed literature on L2 writing assessment. It began with three histories. Yancey's (1999) linear model of trends, where objective measures were followed by timed essays then portfolio assessment, matched New Zealand's history up until 2002. From 2002 to 2007, New Zealand's history matched East's (2008) theory-based model of trends where timed essays and process writing were concurrently used to assess writing. Hamp-Lyons' (2002) prediction of

dynamic online assessment is yet to be fulfilled in New Zealand. As they have been since 2007, the production skills of speaking and writing continue to be assessed by classroom teachers in assessment portfolios. This echoes the importance the MOE places on Assessment *for* Learning theory. The section ended with practical and empirically-tested advice on the implementation of portfolios, noting, in particular, the importance of promoting learning as well as assessment by including in portfolios the sociocultural features such as scaffolding, choice, feedback and reflective practices that have the potential to increase learning. Nevertheless, it was acknowledged that some Assessment *of* Learning principles have persisted during the implementation of the writing assessment portfolios.

The fourth section reviewed literature on Metacognition, Language Learner Strategies and Computer-Assisted Language Learning to establish what has already been said about reflective practices on L2 learning and to find the best topics, language, and modes to guide the sort of reflection that has the potential to increase L2 learning. The literature suggested that learning habits, goals, metalinguistic reflection on teacher feedback, and language learner strategies were suitable topics, that the learner's first language was the most suitable language for L2 beginners, and guided, online diaries (web logs or blogs) were the most suitable modes.

Throughout the literature review, it was difficult not to also notice the number of calls by researchers for more investigations by practitioners, those who, according to Flavell (1979), "would rather assist development than describe and explain it" (p. 909). Wendon (1999) suggested that teachers ask: "What metacognitive knowledge comes into play in the regulation of learning tasks i.e. the use of metacognitive and cognitive strategies, (that are) appropriate to the development of fluency in each of the communication skills i.e. reading, writing, listening, speaking?" (p. 441). Manchón (2011b) called for more longitudinal research and research that is "conducted in real classrooms, with real teachers and real students. Only then will we be able to assess which pedagogical options aimed at promoting learning effectively work in which (foreign language) classrooms and with which students" (p. 59). She also called for research to be conducted on the "levels and types of learning that result from feedback for accuracy and feedback for acquisition" (p. 58). Manchón maintained that the L2 writing fraternity has still to discover which pedagogical interventions are the most likely to foster language learning. Belgrad (2013) called for research which explicitly links reflective writing portfolio processes with achievement and motivation. Murray, Hourigan and Jeanneau (2007) stated that "the main task

for researchers interested in (language learning) will be to improve upon the pedagogical and writing processes involved in turning future 'masterpieces of scattered thought and random concepts' (a quote form a student blog) into better quality academic essays" (p. 27). Ellis and Shintani (2014) noted the lack of research into the effect of using the L1 on L2 learning.

In response to and guided by the literature review conducted in this chapter, and motivated by researchers' calls for more studies to be conducted in real classrooms, on students' reflective processes, strategies and responses to feedback as they write, and on the effective use of the L1 in L2 classrooms, I set out on this doctoral research. In preparation, however, I needed to find researched ways to measure the small, incremental learning gains, and the dynamic and formative movement in interlanguage development that students make in response to daily classroom instruction. The next chapter details a pilot study that I conducted with two classes of my own students. My primary goal was to locate and trial several text-analysis tools that might be capable of objectively measuring L2 learning in the larger study to follow. My secondary goal was to evaluate the potential usefulness of the tools to teachers wishing to measure formative language development for differentiated instruction, especially in portfolio assessment contexts where more subjective and holistic rubrics are the typical tools of measurement.

CHAPTER THREE: FINDING MEASURES OF L2 LEARNING

Following on from the literature review presented in the previous chapter, this chapter details a pilot study that sought to find and trial objective measures of L2 learning capable of providing reliable data for the larger study to follow. It trialled five text-analysis tools which represented five eras in SLA research that were adapted to the classroom context of the study. In the pilot, the tools were employed to analyse texts written in French as Foreign Language (FFL) by four randomly-selected Year 11 students who were preparing for the NCEA Level 1 writing assessment in French. These objective tools were compared for their usefulness and practicality by using the categories of time taken, focus, scope, defining attributes, and benefit to users. One of the tools, frequency analysis, was deemed overly time-consuming and did not supply the type of information useful for classroom research. However, the other four tools: error analysis, obligatory occasion analysis, functional analysis and complexity, accuracy and fluency (CAF) analysis, were found to be potentially useful both for the larger study to follow and also to secondary school foreign language teachers wanting to locate and track learning with a view to differentiated practice and improved learning outcomes for individual students.

3.1: Rationale

Process-based writing portfolios are used to assess the writing of senior students of international languages in New Zealand's standards-based national qualification, the NCEA. To meet the NCEA writing portfolio standard, students must collect, and continue to improve, texts written throughout the year, and then select the pieces that, in their opinion, meet the assessment criteria (see Appendix A). Teachers and moderators judge student achievement against the NCEA writing standard using a holistic rubric of descriptors related to text-type, lexical and grammatical complexity, accuracy and fluency, and ability to communicate information, ideas, and opinions that are unhindered by error (see Appendix B). Students are awarded one of four possible grades according to how well they meet the standard: Not Achieved, Achieved, Achieved with Merit, or Achieved with Excellence.

Because rubrics, such as the ones used for NCEA, judge summative achievement holistically according to a number of language features at the same time, they have practical advantages – they allow researchers, teachers and examiners to grade achievement faster and are therefore more time and cost effective. However, rubrics have two major disadvantages: they require

substantial training and expertise to ensure reliable use (Jonsson & Svingby, 2007), and they are unable to provide the intricate information on micro changes in language learning that researchers need if they are to measure the effects of an intervention, that classroom teachers need if they are to track development and differentiate instruction, and that students need if they are to participate fully in their learning (Moir Scott & Tolosa, 2015).

Hattie and Yates (2013) defined learning as "the process of developing sufficient surface knowledge to then move to deep or conceptual understanding" (p. 26). This movement from surface to deep knowledge requires considerable effort on the part of the learner and the teacher. From a sociocultural perspective, it involves intentionality and reciprocity (Feuerstein & Feuerstein, 1999) where the teacher designs practice activities that are achievable but sufficiently challenging (Vygotsky, 1978) and the learner repeatedly and purposefully engages in those activities and responds to teacher feedback by producing refined output. This reciprocal and intentional process requires both teacher and learner to spend considerable time on "conscious monitoring ... concentration and persistence such that there is stretching to take on new challenges until these challenges become automatic" (Hattie & Yates, 2013, p. 28). However, education stakeholders have tended to value deep knowledge to the detriment of the surface knowledge on which it has been built. This has resulted in a multiplicity of achievement measures yet comparatively few measures of learning (Hattie & Yates, 2013).

This could account for the fact that schools in New Zealand have not always used data to inform teaching and learning as effectively as they might which was signalled by New Zealand's Education Review Office (ERO, 2012). This body (which appraises schools and publishes national reports on current education practice) researched how well schools sought and used data to identify and focus on priority learners who require additional help in order to progress to their full potential. They defined these groups as "ethnic, gender, students with learning needs, students with disabilities and students who are gifted and talented" (p. 11). Of the schools researched, the ERO found that only 9% of schools were highly effective at using data to respond to students' learning needs, 57% of schools were partially effective, and just over one third were either minimally or not effective at all.

Consequently, this small-scale pilot study was motivated to find objective tools for further research and also to address the lack of tools available to classroom teachers wishing to use data

to differentiate practice according to the learning needs of individual students. It trialled and compared five text-analysis tools in the writing units of four Year 11 students of FFL: error analysis, obligatory occasion analysis, frequency analysis, functional analysis, and CAF analysis. It was hoped that these tools, referred to by Hamp-Lyons (1991, p. 248) as "old analytic scoring", might act as complements to the holistic rubrics already used to assess writing in NCEA contexts.

3.2: Learning to Be Measured

For the purposes of this pilot study, language learning is defined according to an interactionist perspective which aligns with Hattie and Yates' (2013) description of learning above. Interactionists view language learning as the discrete changes to a learner's metaphorical interlanguage, and language development as the gradual process of trial and error that leads to fluent use or language acquisition (Norris & Ortega, 2005; Selinker, 1972). Interlanguage changes are made incrementally but are non-linear; they occur as learners receive input and are given ample opportunity for output through interaction with other target language speakers (Ellis, 2005; Fotos & Hinkel, 2007).

In order for interlanguage changes, or learning, to result in interlanguage development and, ultimately, L2 acquisition, Swain and Lapkin (1995) specified three conditions in L2 writing contexts: noticing-triggering, hypothesis-testing and metalinguistic reflection. Noticing-triggering occurs when the learner's attention is drawn to specific forms of language. The student notices the forms and attempts to reproduce them through hypothesis-testing and metalinguistic reflection. These conditions typically result from direct teaching, from carefully-designed practice activities, from Focus-on-Form and from feedback (Hyland, 2011). In communicative contexts, they might also result from self- and peer-review, reflection (or auto-feedback), and the development of learning strategies (Ellis, 2005; Richards, 2006).

Tools capable of measuring formative L2 learning and development rather than assessing summative L2 achievement must therefore be able to track the micro changes that occur in linguistic and cognitive data, such as: becoming aware of and using new language forms or patterns; noticing what has not been previously noticed; attempting to produce language not previously produced; and responding to feedback by restructuring language output (Norris & Ortega, 2005).

3.3: Tools for Measuring Learning

As noted above, the tools selected for trialling in this pilot study were error analysis, obligatory occasion analysis, frequency analysis, functional analysis, and CAF analysis. Their selection was influenced by the language level of the participants, late elementary or A2+ on the CEFR (Council of Europe, 2001), the feedback processes permitted by the New Zealand Qualifications Authority (NZQA, 2006) for use in portfolio writing assessment, the era-specific overview of measures in Ellis and Barkhuizen (2005), and the recommendations regarding CAF measures in the research of Ishikawa (1995) and in the seminal literature review by Wolfe-Quintero, Inagaki and Kim (1998).

Each of the five tools was designed to measure a specific aspect of language and is a pertinent indicator of beliefs about language learning in the era from which it arose (Dyson, 2010; Ellis & Barkhuizen, 2005). By starting with error analysis and ending with CAF analysis, the study journeys through the decades of applied linguistics research and traces how these cognitive tools have moved away from measuring interlanguage development against the target language, towards measuring interlanguage development in terms of the learner (Dyson, 2010). The pilot study addressed three questions: 1. What does the data gathered by each tool indicate about its usefulness for the larger research project to follow? 2. What does it contribute to teacher knowledge of the individual learner and of the individual's learning? 3. Are some tools more useful and practical for classroom practice than others?

The selection of the first tool, error analysis, arose from the process writing standard and one of the types of feedback permitted by the NZQA (2006), indirect, pre-agreed codes in the margins on the lines where errors occur. Error analysis typically involves identifying and describing morpheme errors in learner language. It was primarily used in the 1960s and 1970s to discover how language is learned, but is equally able to show a teacher and learner what has been learned and what is still to be learned (Ellis & Barkhuizen, 2005). Although there has been much debate around the efficacy of error correction (Bitchener, 2008; Bitchener & Storch, 2016; Ellis & Shintani, 2014; Ferris, 2012), error analysis is a valid tool to use in interactionist contexts where errors are an integral feature of interlanguage development and indicate the learner testing out hypotheses, responding to feedback, and discovering language rules in the process (Ellis & Barkhuizen, 2005).

A useful consideration in this context is the distinction made by Corder (1967) and Ellis and Barkhuizen (2005) between error and mistake, errors occurring because of gaps in student knowledge, and mistakes because the student has not yet mastered the form. This distinction is of interest because errors and mistakes give different information about language learning to researchers and they require different responses from teachers. Where errors are likely to require explicit input and further output, mistakes are likely to require indirect input and further output.

The second tool selected for this study was obligatory occasion analysis which focuses on what language learners get right rather than on what they get wrong (Ellis & Barkhuizen, 2005). It ascertains how accurately learners use a specific language form and compares the occasions in which the form is used or attempted to be used with the occasions when the form ought to be used. It was prevalent in the 1970s and 1980s when research focused on determining a natural order to language acquisition (Krashen, 1995) and when the value of instruction was questioned given that natural order (Pica, 1984). It has the potential to provide a more complete picture of an individual's interlanguage development than a focus on only error is able to do.

Frequency analysis, the third tool selected, explores the different linguistic devices used by learners to achieve a specific grammatical form, then computes the frequency of use for each. It was used in the 1980s and 1990s to examine learner language in its own right rather than in comparison with the target language (Ellis & Barkhuizen, 2005). It responded to Bley-Vroman's (1983) comparative fallacy criticism, in which error and obligatory occasion analyses were found wanting because they compared learner language with L1-like expression. Because interlanguage varies from occasion to occasion and between development levels, Bley-Vroman deemed it important to establish patterns of progress by investigating language produced in all its variation and in its own right. A classroom-based frequency analysis therefore has the potential to provide information about the language development patterns of individuals, allowing the practitioner-researcher to design learning activities that are achievable but sufficiently challenging for those individuals (Hattie & Yates, 3013). Frequency analysis recognises language learning as a gradual and variable process and therefore requires longitudinal research to analyse and compare development periods along a timeline (Ellis & Barkhuizen, 2005).

The three tools described thus far (error, obligatory occasion and frequency analyses) all focus on language forms. The fourth tool trialled, functional analysis, arose in the notional-functional

communicative era of the 1980s and 1990s. Most forms of language have a function because they communicate meaning. Functional analysis analyses these communicative functions and the language forms employed to fulfil them (Ellis & Barkhuizen, 2005). For example, the function of 'referring to the future' can be expressed in English by three forms: I *am going to work* harder, I *will work* harder, and I *shall work* harder. In mapping the relationship between functions and forms, functional analysis has the potential to provide information regarding the developing richness of an individual's lexis and syntax.

The fifth tool selected for trial, CAF analysis, arose from beliefs about language learning typical of the 1990s and 2000s. In this era, many previously popular notional-functional syllabi developed into standards-based national curricula influenced by an exponential increase in the use of educational technologies (Sewell, 2011) and Assessment *for* Learning practices (I. Lee, 2007). CAF analysis assumes that as language learners progress, they will: write sentences that are more grammatically and lexically complex, write with fewer errors, and write more words in a given time period (Wolfe-Quintero, et al., 1998). While CAF analysis measures complexity, accuracy and fluency separately, it has led to the development of the currently-prevalent, time-efficient rubrics which describe these features separately but score them holistically (I. Lee, 2007). Such rubrics are used to measure achievement in New Zealand's national qualification, the NCEA.

Table 3: The Most Effective CAF Measures (Wolfe-Quintero, et al., 1998)

Development in:	Most Effective Units of Measurement
Grammatical Complexity	C/T, DC/C
Lexical Complexity	WT/√2W, SWT/WT
Accuracy	EFT/T, E/T
Fluency	W/T, W/C, W/EFT

In their seminal literature survey on what they called *objective development measures for second language writing*, Wolfe-Quintero, et al. (1998) investigated thirty-nine research projects, comparing more than one hundred CAF measures. As indicated in Table 3, they concluded that the best measures of development in grammatical complexity were the ratio of clauses per T-unit (C/T) or dependent clauses per clause (DC/C); the best measures of development in accuracy were the ratio of error-free T-units[9] per T-unit (EFT/T) or errors per T-unit (E/T); and the best measures of development in fluency were the ratio of words per T-unit (W/T), words per clause

---

[9] A T-unit is defined as a main clause and the subordinate clauses connected to it.

(W/C) or words per error-free T-unit (W/EFT). Wolfe-Quintero, et al. suggested that, in addition to research purposes, these measures could "aid classroom teachers in answering small-scale research questions within the context of their own classrooms" (p. 126).

It is important to note that the participants in the vast majority of projects reviewed by Wolfe-Quintero, et al. (1998) were young ESL adults in university settings. Because CAF analysis is a situated measurement tool (Norris & Ortega, 2005), researchers must choose measures that suit their research context. Ishikawa (1995) established the clause rather than the T-unit as the more suitable measure of development in low proficiency data. Using a battery of twenty-four measures, she scored the writing of low-proficiency ESL learners and found the measures that best discriminated between low and close levels of language were the total words in error-free clauses or the number of error-free clauses per composition. Also referring to beginner levels of language, Ellis and Barkhuizen (2005) suggested supplementing the measurement of language complexity by subordination with study of a particular language feature, such as verbs.

These findings are particularly relevant in the context of this pilot study and the larger study to follow because most Level 1 NCEA candidates have been learning secondary school French for two or three years only and many are not yet writing at an intermediate level, so their language production is not complex. CAF measures that use clauses are, therefore, more suitable. In addition, the writing standard requires students to refer to past, present, and future experiences, so they learn the four foundational French verb tenses, present, past perfect, past imperfect, and future, that allow them to fulfil this requirement. A study of verb use is, therefore, also appropriate.

To sum up, this first section of the chapter has introduced the rationale for the pilot study, established the learning to be measured, and the measurement tools to be trialled: error, obligatory occasion, frequency, functional and CAF analyses. The goal of the pilot was to ascertain the suitability of each tool for gathering language learning data for the larger research project to follow. In addition, it aimed to ascertain the suitability of each tool for gathering the kind of data that the ERO (2012) envisaged would inform teaching practice, and the kind of data that Hattie and Yates (2013) envisaged would make daily learning visible. Each tool was designed to measure a specific aspect of language learning and was a pertinent indicator of beliefs held by SLA researchers in the era from which it arose (Dyson, 2010; Ellis & Barkhuizen, 2005). As these text-analysis tools are arguably more consistently reliable than rubrics (Hamp-Lyons, 1991), they

are potential tools for use by classroom researchers who wish to strengthen reliability of data as well as by classroom teachers who wish to collect data on learning in addition to the holistic achievement data that national assessment rubrics provide.

3.4: Pilot Study

As previously stated, this pilot study addressed three questions: 1. Are the selected tools suitable for gathering L2 learning data for the larger research project to follow? 2. How does the data gathered by each tool contribute to the teacher's assessment of individual student's L2 learning? and 3. Are some tools more useful and practical for classroom use than others?

3.4.1: Participants

The study collected writing samples from four participants chosen randomly from a Year 11 class of fifteen 14 to 16 year old girls taught by the practitioner-researcher. Following ethics protocols, the students were randomly selected using a school-wide computerised process, approved by the examining body, the NZQA. When the study was implemented, the students were beginning their third year of secondary school French which meant they were writing at a pre-intermediate level, or A2+ on the CEFR (Council of Europe, 2001).

During the first term of the academic year, all students in the class wrote a set of texts in French containing a pre-test online profile, a first and second draft of two health-related dialogues, and a post-test sports and leisure survey. In the classroom where the participants produced these samples, process writing, portfolio assessment, reflective journals, strategy use and goal setting were important components of the programme.

3.4.2: Language Samples

For four of the text-analysis trials, the language samples were the same text-type (two health-related dialogues) to increase reliability (Benevento & Storch, 2011; Kuiken & Vedder, 2008; Nunan & Bailey, 2009; Schoonen, et al., 2009). To trial frequency analysis, however, the pre- and post-tests were also included. The pre-test web profile was written under test conditions, like the dialogues, but the post-test, a survey, was written for homework. All were gathered during the usual learning sequence and extended the research period from six weeks to three months. The language samples were produced in response to specially designed tasks so they were not considered naturalistic or authentic. However, they met Ellis and Barkhuizen's (2005) definition for

clinically elicited samples which have "psycholinguistic validity in that the conditions of language production (e.g. a primary concern for message conveyance and with incidental attention to form) are similar to those found in naturally occurring language use" (p. 48). To establish reliability in the analysis of the errors, a L1 French colleague marked the two first drafts and the post-tests of each of the four participants. Error agreement between the researcher and the colleague was calculated and an absolute agreement of 95% was reached.

Table 4: Pilot Study Process Writing Sequence

Week 1	Writing task and assessment schedule were distributed one week in advance to facilitate student planning
Week 2	50 minute Draft 1 was written under test conditions with teacher as resource for general questions
Week 3	Teacher provided coded feedback on first draft
	Students tallied codes to find their most frequent errors
	Full class teaching on common errors was conducted
	Individually, students corrected their most obvious errors
	Students were directed to language practice exercises to aid error correction and rewriting
Week 4	Students reflected in blogs
	Students rewrote for homework and submitted second draft within a week

The process writing unit that elicited the primary writing data (two drafts for each of two tasks) arose from the Health and Well-being topic being taught at the time. The students were given the instructions a week in advance (see Table 4) to facilitate planning at home and were allotted fifty minutes in class time under test conditions and without resources to compose the first draft. This allowed teacher answers to benefit all students while minimising interruption to writing flow. In a later class, the students were allotted fifteen minutes to consider teacher feedback which was offered in three ways according to the NZQA (2006) clarifications: pre-agreed codes in the margin of lines where language errors occurred, global comments to encourage and indicate steps that might be taken to improve the script against the standard, and full class teaching on common errors. On this occasion, the students were asked to tally their own error codes and find their most frequent errors. If they were unable to correct any errors, they were directed by the code schedule to a choice of specific practice activities. These steps prepared them for the second draft, which they wrote for homework. After the first and second drafts from the two process writing units were collated, the students received an holistic NCEA-type grade to indicate progress towards the writing standard (see Appendix B). At this point, the four participants were randomly selected from the class and their samples analysed using the five tools: error analysis, obligatory occasion analysis, frequency analysis, functional analysis, and CAF analysis.

3.4.3: Data Analysis

For the error analysis, learner error was coded according to broad and general categories using terms meaningful to the adolescent learners, such as, *V* for verb error, *Sp* for spelling error and *U* for an idiom or usage error (see Appendix F). The broad nature of the codes helped to minimise problems that typically arise in error analysis around the definition of error and around counting errors when multiple error boundaries overlap. Error severity has also been a concern for researchers (Evans, et al., 2014) and several have devised and trialled three and four level scales to measure error based on degrees of communicative seriousness (Foster, et al., 2000; Kuiken & Vedder, 2008). Such scales answer questions posed by linguists who require standalone, generalisable measures but using the communication-based NCEA rubric in conjunction with the error analysis minimised such a concern in this study.

For the obligatory occasion analysis, article, noun and adjective agreements were tracked in the first drafts of the two dialogues, agreements being a feature of French language with which elementary (A2+) students often struggle. Obligatory occasions and correct attempts were tallied, and a percentage of accuracy calculated. Then, following Corder (1967) and Ellis and Barkhuizen (2005), all inaccurate occasions were categorised as either error or mistake based on the practitioner-researcher's knowledge of the curriculum and the course, and whether the language form was produced accurately or inaccurately in most instances (see Section 2.2.3).

For the frequency analysis, progress using the various forms of the partitive article (*de, du, de l', de la, des*) was examined in the first drafts of the two dialogues and the pre- and post-tests. This was another language feature often wrongly applied by students at this level. The three months' collection period was divided into four time periods delineated by the production of the four first drafts. It located and recorded all uses of the partitive article, correct and incorrect, and calculated the percentage of correct uses over total use in each time period.

The functional analysis examined just the two first drafts of the Health and Wellbeing unit in which dialogues between patients and medical professionals were written, the first to gain a health certificate for a travel visa, and the second to explain and take advice on an illness or injury. The number and timing of language forms used to fulfil the function of 'making future plans' were investigated.

Finally, CAF analysis was used to measure language learning in the first drafts of the two dialogues. The language level of the participants was considered when deciding which CAF measures to trial. Complexity was measured with the percentage of dependent clauses per clause (DC/C), with students tallying the number of verb tenses they employed as a consciousness-raising activity. To measure accuracy, the study used the percentage of error-free clauses per clause (EFC/C). Because the first drafts were written under timed test conditions, it measured fluency by total words written in fifty minutes (W/Comp). By using all three types of measure with one set of learners, this CAF analysis attempted to fill a research gap highlighted by Wolfe-Quintero, et al. (1998).

3.5: Results and Discussion

The data from the trial of the five text-analysis tools were analysed firstly, tool by tool and student by student, and then, secondly, by comparing the five tools. In this section, the results from the tool by tool and student by student analysis are presented.

3.5.1: Error Analysis

Error analysis involves identifying and describing errors in learner language. In the error analysis of the eight first drafts (two drafts per participant) seen in Table 5, by far the most consistently frequent error type (25 errors) was the articles and particles that do not always have a semantic and/or grammatical equivalent in English, e.g. Student 1 wrote: *J'essaye manger beaucoup de fruits* instead of *J'essaye de manger beaucoup de fruits*. (I try to eat lots of fruit.) The next most frequent error types were spelling (13 errors) which included accents and meaning-significant punctuation, e.g. Student 2 wrote: *Je mange équilibre* instead of *Je mange équilibré* (I eat a balanced diet), and nominal agreements (13 errors), e.g. Student 4 wrote: *Je n'aime pas le oignon* instead of *Je n'aime pas les oignons* (I don't like onions). These errors were closely followed in frequency by errors in verbs (12 errors), e.g. Student 3 wrote: *Vous ne veux pas vous rapprocher* instead of *Vous ne devez pas vous rapprocher* (You shouldn't get too close to me).

In six out of eight second drafts, students corrected an average of 84% of errors. In the remaining two second drafts, the correction rates were 56% and 50%. Over all eight second drafts, the errors not corrected included seven agreements, six linking words, and one verb. These errors, three of which were changed wrongly, indicated a need for further input. A few new errors also

appeared in the rewriting. However, in the majority of instances, the indirect error codes and the activities built around those errors were successful in helping participants improve their accuracy.

Table 5: Pilot Study Error Analysis

Student (S) and Task		Most Frequent Errors in Task 1 Draft 1		Response to Codes Task 1 Draft 2	Task 1 Errors Repeated in Task 2	Next Steps
S1	Task 1	Linking words Omissions	7 6	16/22 (73%) corrected 5 Linking errors remained	• Very few errors in second task	• Practice on verbs that link with *de* to infinitive
	Task 2	Verbs Omissions	3 2	6/6 (100%) corrected	• Linking words much improved • Two omissions but typos?	
S2	Task 1	Linking words Spelling Verbs	5 3 3	19/20 (95%) corrected 1 omission remained	• Errors of spelling (acute accents and hyphens) repeated in second task	• Practice on agreements
	Task 2	Spelling Verbs Agreements	7 6 5	16/20 (80%) corrected 3 agreement errors remained 1 verb error remained		
S3	Task 1	Agreements Linking words	6 6	13/17 (76%) corrected 4 agreement errors remained	• First task linking errors repeated in second task	• More practice on partitives and agreements
	Task 2	Linking words Spelling	3 3	5/10 (50%) errors corrected but 4 new ones occurred in Draft 2 (typos)	• 4 agreement errors not corrected	
S4	Task 1	Register Agreements	3 2	5/9 (56%) corrected 2 errors wrongly correctly: one negative, one possessive adjective Two new typos? in draft two	• First task errors not repeated in second task.	• Register appropriate to text-type to be made more explicit
	Task 2	Linking words	4	4/5 (80%) corrected 1 article error corrected wrongly		

As indicated in Table 5, for two of the participants at least, Student 1 and Student 4, the error analysis and the teaching and learning sequence that surrounded it appeared to have aided language learning as indicated by increased accuracy from the first to the second task. For these two students, errors made in the first task were greatly reduced or eliminated in the second task, although new types of error did occur. These new errors indicated a need for more output that included these particular language features. However, Student 2 repeated spelling errors in the second task in the form of accents and hyphens and Student 3 repeated errors in partitive articles (a subset of the errors listed as 'linking words'). In this instance, these students required more explicit input than the error codes offered.

There has been much discussion about the efficacy of error correction (see Section 2.2.3). Dyson (2010) stated that error correction, "by focusing on the errors ... emphasises transition to the target language and suggests linguistic failure" (p. 33). Nevertheless, the situated error analysis used in this study proved valuable in locating the L2 learning of individuals and in providing those individuals with opportunities for taking initiative in their own learning. They counted their own error codes to find their two or three most frequent errors, then chose from, and took part in, a range of practice activities before rewriting. By inviting student participation in the error-analysis sequence, coded feedback has the potential to initiate the type of deep linguistic processing that promotes language learning (Ferris, 2006; Manchón, 2011a).

3.5.2: Obligatory Occasion Analysis

Obligatory occasion analysis measures how accurately learners use a specific language form and compares the occasions in which the form is used or attempted to be used with occasions when it ought to be used (see Table 6). In this pilot study the occasions when gender and number agreement (another frequent error type) were required were tallied in each of the eight first drafts (two first drafts per participant) and the percentages of appropriate use were calculated. On the occasions of inappropriate use, a distinction was made between error and mistake based on the practitioner-researcher's knowledge of the curriculum and the course, and whether the form was produced mainly correctly or incorrectly throughout the script.

Table 6: Pilot Study Obligatory Occasion Analysis

Student and Task		Tally of Agreement Occasions	Percentage of Appropriate Agreements	Error or Mistake?	Changes Task 1 - 2	Teaching Practice Implications
S1	Task 1	17	17/17 - 100%	0	Similar	
	Task 2	44	42/44 - 95%	2 mistakes		
S2	Task 1	78	75/78 - 96%	2 mistakes 1 error	Similar	Find teaching moments in existing programme for:
	Task 2	32	28/32 - 88%	2 mistakes 2 errors		• tout in all its forms • se sentir + adjective • le corps is singular • quel in all its forms
S3	Task 1	49	43/49 - 88%	1 error 5 mistakes	Similar	Create a checklist for proofreading?
	Task 2	13	13/13 - 100%	0		
S4	Task 1	72	71/72 - 99%	1 mistake	Similar	
	Task 2	30	30/30 - 100%	0		

When calculating across all eight first drafts, it was found that there were 335 obligatory occasions, in 96% of which agreement conventions were applied appropriately. Of the sixteen

occasions when conventions were inappropriately or not applied, only four appeared to be errors, the remaining appearing to be mistakes. When comparing the growth of knowledge between the first drafts of Task 1 and Task 2, there were not enough differences to ascertain whether language learning had occurred or not. Instead, contrary to teacher belief, it appeared that, at this pre-intermediate level of French, these four students had already mastered nominal agreements. The errors (as compared with mistakes) that did occur, occurred in aspects that textbooks and/or teachers had not yet introduced: the various forms of *tout* (all/every), subject agreement after *se sentir* (to feel), the interrogative adjective *quel* (which), and the singular status of *corps* (body) (see Table 6).

Being not overly time-consuming (just under four minutes to locate data in Student 1's Draft 1 Task 1), obligatory occasion analysis is a valuable tool that enables both researcher and teacher to ascertain how much is known and used well by participants as well as how much is not known and is still to be learned. It provides balanced, whole-picture data on one language feature and thereby has the potential to rectify incomplete perceptions regarding that feature. When incorrect occasions are categorised as error or mistake, next steps in the learning sequence are clarified.

3.5.3: Frequency Analysis

Frequency analysis explores the different types of linguistic 'device' used by learners to achieve a specific grammatical form, then computes the frequency of use for each 'device'. To trial frequency analysis in this pilot study, the partitive article was examined in the four first drafts gathered from each participant over three months (sixteen drafts in all). As all attempts to form the partitive were recorded as well as the correctness of these attempts, the amount of data gathered was extensive. Table 7 is a summary of these data and indicates the number of correct attempts per attempt for each of the four tasks per participant over the four time periods. From the full data not reported here, it was evident that all students experienced difficulty using the partitive article in instances when it has no equivalent in English. They all showed inconsistent use across the four pieces of writing.

Table 7: Pilot Study Frequency Analysis – Correct Usage per Usage over Time

	Pre-Task		Task 1		Task 2		Post-Task	
S1	5/6	83%	11/15	73%	9/9	100%	27/31	87%
S2	5/9	56%	24/30	80%	7/8	88%	22/26	85%
S3	12/16	75%	6/9	67%	3/5	60%	10/14	71%
S4	6/8	75%	16/16	100%	6/10	60%	2/6	33%

However, Students 1 and 2 regularly used some of the more complex aspects of the partitive article for students of this level, e.g. *de bonnes qualités* (good qualities) and *d'autres langues* (other languages) (rule: the partitive *des* becomes *de* before a preceding adjective). Student 3 mastered the use of *de* to indicate possession and *de* in expressions of quantity, even though one in three or four uses of the partitive remained faulty. In Task 1, Student 4 wrote sixteen perfectly formed partitives, yet the other three scripts contained poorly formed versions. Even so, despite continued errors, some complex understandings were evident in this student's work, e.g. *les symptômes d'un rhume (the symptoms of a cold), prenez de l'aspirine* (take some aspirin)*, avez-vous d'autres conseils (do you have any other advice).*

3.5.4: Functional Analysis

Functional analysis focuses on language functions and the various language forms employed to fulfil them. It was used in this pilot study to locate and track the forms that participants used to fulfil the function 'making future plans'. An analysis of the first drafts of the two dialogues highlighted the emergence of the future tense, which had been introduced in the three weeks between the two writing units. Prior to this, students used the near future (*aller* + *infinitive*), for example, *je vais manger* (I am going to eat). Student 1 illustrated understanding of the near future with one correct instance in Task 1 and uptake of the simple future by using a correctly formed irregular verb and reflexive verb in Task 2 (see Table 8).

Table 8: Pilot Study Functional Analysis

Student/Task		Forms Used to Refer to Future Plans	How Informs Teaching Practice
S1	Task 1	Je vais aller	Correctly formed irregular and reflexive future verbs
	Task 2	Je reviendrai, vous vous sentirez mieux	
S2	Task 1	Quand allez-vous aller, je vais voyager, vous allez recevoir	Lack of confidence with the newly introduced form?
	Task 2	None	
S3	Task 1	None	Must show more of what is known
	Task 2	Vous vous sentirez mieux	
S4	Task 1	Je vais voyager	Correctly formed negative, regular and reflexive future verbs
	Task 2	Vous n'allez pas mourir, je suivrai, vous vous sentirez mieux	

Student 2 used a range of near future verb combinations in Task 1, but avoided any reference to the future in Task 2. Student 3 did not use any near future forms in either task, and only one simple future, albeit with a faultless reflexive verb, in Task 2. Student 4 used a near future in

Task 1, a negative near future and two simple futures in Task 2, one with a reflexive verb. All but Student 2 showed emergence of the simple future, and all but Student 3 showed familiarity with the near future. This time-efficient tool (1 ½ minutes for Student 1's Draft 1 Task 1) indicated what each student needed to include in subsequent pieces of writing if the NCEA requirement to refer to past, present and future occasions was to be fulfilled. It also indicated to the practitioner-researcher the individual students' uptake of the previously taught and recently taught forms.

3.5.5: Complexity, Accuracy and Fluency Analysis

CAF analysis measures complexity, accuracy and fluency in writing arriving at three separate scores (see Table 9). Given the language level of the research participants and based on the findings of Wolfe-Quintero, et al. (1998) and Ishikawa (1995), the study measured complexity by dependent clauses per clause (DC/C), accuracy by error-free clauses per clause (EFC/C) and fluency by the number of words written in a fifty minute time period (W/Comp).

When comparing the first drafts of the two tasks for increase in complexity, two of the four students produced an increased percentage of dependent clauses per clause; when comparing the first drafts of both tasks for growth in accuracy, three of the four students produced an increased percentage of error-free clauses per clause; and when comparing the first drafts of both tasks for fluency, none of the students increased the number of words per composition.

Table 9: Pilot Study CAF Analysis

Student and Task	Complexity by % DC/C		Accuracy by EFC/C		Fluency by W/Comp
S1 Task 1	4/73	5.5%	57/73	78.0%	417
Task 2	7/56	12.5%	49/56	87.5%	290
S2 Task 1	5/69	7.2%	55/69	79.7%	476
Task 2	2/42	4.7%	25/42	55.5%	236
S3 Task 1	7/44	15.9%	32/44	72.8%	266
Task 2	2/32	6.3%	24/32	75.0%	157
S4 Task 1	5/66	7.6%	57/66	86.4%	345
Task 2	6/40	15.0%	36/40	90.0%	219

It appears that measuring fluency by the number of words written in a timed period was not an appropriate tool for the context. By this measure, all four participants showed a considerable drop in fluency in their second task. Although both tasks were dialogues on a Health and Wellbeing topic being taught at the time and were set and written under the same conditions, students were clearly not able to produce as much content in the second task. In hindsight, it was clear that,

because quality rather than quantity is a direction of the NCEA writing standard, measuring fluency using the quantity of words written in the time allocated does not necessarily indicate the level of fluency in the text. For instance, Student 4, whose work was the most complex and accurate, would be considered by this measure to be the least fluent. Instead, she was the most skilled at showing the quality of her language to meet the standard using the least number of words. This discrepancy highlights comment by Norris and Ortega (2005) that "[t]here is certainly much more to language learning and language use than what measures of CAF might account for. We probably need to be careful in our zeal for focusing on CAF that we do not ignore other phenomena essential to a more complete understanding of second language learning" (p. 575). Perhaps using the number of error-free clauses per composition, also found by Ishikawa (1995) to be effective for low proficiency scripts, would have been a better measure of fluency in this instance.

3.6: Comparing the Five Tools

After looking at each text-analysis tool student by student and considering the information it provided, the five tools were compared for usefulness, and practicality using the categories of time taken, focus, scope, defining attributes and benefit to users. When comparing time taken, each of the five tools gathered data from Student 1's first draft of the first dialogue only (see Table 10). CAF analysis proved the most time-consuming tool (6m 26s) with DC/C taking 2m 28s, EFC/C 38s, and number of words taking 3m 20s. Error analysis took two minutes less (4m 30s), obligatory occasion analysis and frequency analysis took similar amounts of time (3m 50s and 3m 40s), and functional analysis proved the most time-efficient (1m 25s).

Table 10: Pilot Study – Time Taken To Locate Learning in Student 1's Draft 1 Task 1

Tool	No. Drafts	No. Tasks	No. Minutes	Total Time
Error	2	2	4m 30s	18m
Obligatory Occasion	1	2	3m 50s	7m 40s
Frequency	1	4	3m 40s	14m 40s
Functional	1	2	1m 25s	2m 50s
CAF	1	2	6m 26s	12m 52s
NCEA Rubric	1	2	2m 20s	4m 40s

m=minutes, s=seconds

However, to ascertain language learning, the study needed to compare more than one script per student (see Table 11). Error analysis used the first and second drafts of both dialogues (four scripts per student), frequency analysis used four first drafts per student, obligatory occasion and functional analysis used two first drafts per student, and CAF analysis used three measures to

gather data from two first drafts per student. After timing and comparing the time taken for each of the five text-analysis tools, the time taken by each tool was then compared with the time taken to grade the same script using the mandated NCEA rubric. Although designed to measure a portfolio of several tasks, when used to assess just one piece of writing this holistic tool took less time than four of the five tools trialled. In addition, grading with a rubric can take place simultaneously with other analyses to render it even more time-efficient.

As well as time taken, each tool was compared using metaphorical photography terms, according to focus and scope (see Table 11). Whereas error analysis focused on all errors in all forms, providing a wide and negative view of learning, obligatory occasion analysis focused on one form, viewing it from both positive and negative angles. Frequency analysis focused on the close up view of the details of one form, while functional analysis focused on multiple forms used to perform one function, thus providing a shallow but wider view of language learning. The focus of CAF analysis was on three language features which together provided a more panoramic view of learning.

Table 11: Pilot Study Comparison of the Five Tools

Analyses	Error	Obligatory Occasion	Frequency	Functional	CAF
Time	4m 30s (x4)	3m 50s (x2)	3m 40s (x4)	1m 36s (x2)	6m 26s (x2)
Focus	multiple forms	one form	multiple devices for one form	multiple forms of one function	three language features
Scope	wide but negative view	narrow but negative and positive views	narrow but detailed view	narrow and close up view	wide and complex view
Attribute	NCEA-approved feedback integral to process writing	balanced	longitudinal	useful for tracking emerging forms	multiple measures to use alone or in combination, yet still incomplete
Benefit to whom	R, T, S	R, T	R	R, T, S	R, T, S

m = minutes, s = seconds, T = teacher, S = Student, R = Researchers

Finally, the study compared the particular attributes that defined each tool and looked at each tool's usefulness and to which users (see Table 11). Error analysis was non-negotiable in this study as feedback on error in the form of indirect codes is an integral and expected component of the NCEA-approved portfolio assessment process. The codes proved useful to the practitioner-

teacher and the students as they facilitated feedback for acquisition as well as feedback for accuracy (Manchón, 2011b; Norris & Ortega, 2005). Feedback for acquisition in the pilot study was achieved via student involvement in the data gathering process and in the choice of remedial practice activities that were offered between the writing of the first and second drafts. Feedback for acquisition was also achieved by providing students with time to reflect in journals on those choices and to plan their next steps. Obligatory occasion analysis provided the practitioner-researcher with balanced data on students' correct and incorrect use of one form. Contrary to expectation, it highlighted the fact that these students had already mastered nominal agreements. It also highlighted the next steps to be taken in the teaching and learning sequence. Frequency analysis required a longitudinal study more suitable for research than classroom teaching and learning because of the heavy time commitment and the amount of detail collected which was unrealistic and not useful to either teacher or students in the classroom context. Functional analysis proved time-efficient and effective for researcher, teacher and student alike because it located and tracked an emergent form and made evident what had been learned and was still to be learned. The complexity and accuracy measures used in the pilot were also potentially useful to researcher, teacher and student while the fluency measure was not. A better measure of fluency might have been the number of error-free clauses per composition, as recommended by Ishikawa (1995). An additional complexity measure, the number of different verb tenses produced, although not used by the practitioner-researcher in the pilot study, was used by the students to gather information regarding the tenses they needed to include in subsequent pieces of writing. (If they were to meet the specification of the NCEA writing standard they needed to show they were able to refer with consistency to past, present and future events). The study found that, when CAF analysis is used in conjunction with a rubric containing a communication component, it is a most useful tool for tracking whole picture language learning and development over time and therefore was the tool chosen for use, in conjunction with error analysis, in the larger study to follow.

There are two features of this pilot study that limit generalisation: the short length of time in which the primary data collection took place (six weeks) and the small number of participants (four). However, as a pilot study designed to provide information regarding appropriate text-analysis tools for measuring language learning in the larger study to follow, it offered much insight. From a practitioner's point of view, the information gained from the analysis of the tools was useful in

understanding whether or how each tool (that had primarily and previously been used in research) might work in the NCEA classroom context.

3.7: Summarising the Chapter

The aim of this pilot study was to determine which of the five selected text-analysis tools might prove effective and useful complements to the holistic rubrics used in New Zealand's standards-based national assessments, that is, whether they might effectively locate and track L2 learning in the writing of four Year 11 FFL students, whether they might be useful to the larger research project to follow, and whether they might inform classroom teaching and learning. The research responded to Manchón's (2011a) call for research to be conducted by real teachers in real classrooms in order to discover what promotes language learning and in which contexts; it responded equally to Hattie and Yates (2013) who claimed that there is a proliferation of achievement measures but few measures of learning; and it responded to the ERO (2012) who found that only 9% of New Zealand schools were highly effective in locating and tracking learning for improved outcomes.

Of the tools trialled for practitioner research, obligatory occasion analysis proved effective in highlighting the correct as well as the faulty applications of one language form; frequency analysis proved too time consuming for classroom use and did not supply the type of information useful to daily teaching and learning; functional analysis proved worthwhile for tracing the emergence of a new form; and two measures used in the CAF analysis, the percentage of dependent clauses per clause to measure complexity, and the percentage of error-free clauses per clause to measure accuracy, proved useful complements to holistic rubrics. However, the tool used to measure fluency, the number of words written in a fifty minute period, was found to be without data validity as it did not measure what it set out to measure. The tool that proved the most valuable was the situated form of error analysis employed. It was valuable for three reasons: it was embedded in the process writing sequence and used broad codes understood by learners to describe their error types; it prepared learners to write an improved second draft; and it invited them to be involved in their own feedback process by tallying their own errors and by choosing from a range of practice activities designed to inform error correction.

In conclusion, no single tool exists that will effectively measure language development (Moir Scott & Tolosa, 2015). Four of the text-analysis tools trialled proved useful for the secondary school

foreign language teacher who wants a more fine-grained analysis of second language development in writing and who may want to adapt error, obligatory occasion, functional and CAF analyses to their classroom contexts in order to inform their teaching and learning programmes. For the researcher wishing to find objective tools to measure learning in L2 writing, complexity and accuracy measures in combination with error analysis and holistic rubrics were found to be the most useful tools. For the purposes of the larger research study to follow, the ratio of dependent clauses per clause and of error-free clauses per clause were deemed to be appropriate measures of complexity and accuracy in combination with a more holistic rubric. The number of words written in a timed period was deemed an inappropriate tool for measuring fluency in the research context. Instead, the number of error-free clauses per composition could produce more valid data (Ishikawa, 1995).

In this chapter, the pilot study established the most appropriate tools for locating L2 learning gains in the writing of the adolescents who participated in the larger study to follow. The next chapter turns from the pilot study to the larger study and considers, in addition to measuring tools, an appropriate research design. It begins by exploring three studies conducted in contexts very similar to the present study, in the hope that they might provide direction for a design that will produce robust evidence to address the research questions posed.

CHAPTER FOUR: CHOOSING A RESEARCH DESIGN

SLA researchers have long been preoccupied with describing L2 acquisition and discovering what promotes it. However, depending on their beliefs about SLA theory, they define and measure acquisition very differently. Norris and Ortega (2005) specified six steps to conducting research that will result in an understanding, if not an acceptance, of findings by those holding different theoretical beliefs (Norris & Ortega, 2005). The first and second steps involve defining the constructs to be measured and identifying the behaviours to be observed. The third and fourth steps involve specifying the tasks that will elicit such behaviours and then eliciting, recording and observing the behaviours while controlling or accounting for variables. The fifth and sixth steps consist of scoring the observed behaviours in a way that connects them to interpretations, and summarising and analysing the data according to probable categories.

This chapter is divided into three sections. The first section explores the ontological beliefs that led to the selection of a research methodology. The second section examines two previous interactionist intervention studies and one sociocultural intervention study that have informed the present research design. The third section outlines the mixed method, counter-balanced, comparative study that is presented in this thesis. It follows Norris and Ortega's (2005) six steps and uses both interactionist and sociocultural tools to measure the language learning evident in the participants' L2 writing when they reflect and when they do not.

The Year 11 participants (n = 71) came from five classes in four New Zealand coeducational secondary schools where they were learning French at Levels 5 and 6 of the *Learning Languages* curriculum of the NZC, which equates with Global Scale Levels A2 and A2+ of the CEFR for Languages (Council of Europe, 2001). The participants for the research were able to include the writing they produced in their writing portfolios and to present it for assessment against the NCEA Level 1 writing standard.

The study set out to address the following questions:

1. Does L1 written reflection increase L2 learning when adolescents write for portfolio assessment? If so, what is the type and amount of increase and for whom?
2. What insights, if any, does the content of the L1 reflections reveal concerning the results of the L2 data analysis?

3. Are there any links between the reflective self-regulatory practices reported in the reflections, achievement levels, and/or continued learning?

4. What further insights regarding language learning do the reflections provide?

The chapter begins by exploring the ontological beliefs that led to the selection of the research methodology.

4.1: Research Paradigm, SLA Theory and Research Design

When interpreting empirical research for application in their own classrooms, Nunan and Bailey (2009) recommended that teachers use experience and common sense while Cohen, Manion, and Morrison (2007) referred to experience and reasoning. Both sets of researchers agreed that we make sense of our world in three distinct ways: through our experience, our reasoning and our research. Where experience and reasoning may result in an incomplete view of reality, research enlightens experience and reasoning with systematic investigation (Cohen, Manion, & Morrison, 2007).

Nevertheless, there is more than one way to view reality. At one end of a continuum, researchers view reality as static and finite, and with measurable universal truths (the positivist stance). At the other end, they view reality as a dynamic creation of the human mind with multiple contextualised truths (the constructivist stance). In between are situated combinations of these two extremes. The position that researchers take on this continuum (their ontology) inevitably frames the design of the research they conduct; this is their research paradigm.

In a review of research conducted between 2001 and 2006, Harklau and Williams (2010) found that L2 writing researchers do not often report their research paradigm, nor do they explicitly state their beliefs about SLA. However, no matter how understated the researchers' beliefs are, their choice of paradigm is most often visible in their choice of research design. For example, Nunan and Bailey (2009) explained that, in the behaviourist 1960s, SLA research took a positivist position on the paradigm continuum and framed their research in Generative Linguistics theory. Researchers typically sought to uncover what was universal about teaching and learning by using experimental research methodologies that collected quantitative data leading to generalisable findings. Classroom research typically involved teaching two groups of randomly assigned students using two different methodologies and measuring the learning outcomes to find the

difference, thereby establishing the impact of the variables under investigation. To minimise interference, the researchers attempted to control all other variables.

In the 1970s and 1980s, due to the flourishing of interactionist theories and CLT methodologies, SLA research edged away from the positivist end of the paradigm continuum and closer towards the constructivist end. The researcher became more subjective, used methodologies more concerned with meaning than accuracy, and collected qualitative rather than quantitative data (Cohen, et al., 2007). Nunan and Bailey (2009) described such research as post-positivist and naturalistic because it sought to gain insight into the complexities of language teaching and learning in uncontrolled classroom contexts. In pure versions of naturalistic research, researchers typically used large-scale longitudinal ethnographies or in-depth case-studies to gather data that is concerned with the qualities and attributes of learners rather than with measuring and counting aspects of their language learning (Nunan & Bailey, 2009).

In the 1990s and 2000s, sociocultural theory built on interactionism and influenced both research and classroom practice. Typical SLA research conducted in this era nudged still closer to the constructivist end of the paradigm continuum. The methodologies used by researchers combined both positivist and post-positivist features. However, the knowledge gained about language learning was interpreted by very different thinking. Lantolf and Beckett (2009, p. 459) listed the "Zone of Proximal Development, private speech, internalisation, regulation, activity theory, and the genetic method" as features of sociocultural research in this period.

Because some aspects of language acquisition are universal while others are contextual, positivist and constructivist research paradigms are not necessarily mutually exclusive; SLA research can be concurrently pure and applied and normative and interpretive (Ellis, 2005). In order to distinguish the universal from the contextual, SLA researchers use mixed methods to such an extent that mixed method research could be considered a paradigm in its own right (Ellis & Barkhuizen, 2005). Nunan and Bailey (2009) specified mixed methods as a guiding principle for SLA research and added that it is not uncommon for an L2 researcher to adopt a normative research design, gather qualitative data with quantitative collection methods and analyse the data qualitatively. In fact, Johnson (1993) went so far as to say that "being bimethodological or multimethodological is a mark of scholarly sophistication," and that "in the field of Second/Foreign

Language learning and teaching as a whole… a more multimethodological emphasis is essential"
(p. 4).

It seems, therefore, that the popularity of mixed method research in the field of SLA indicates a
strong link between SLA theory and research design and a looser link between research
paradigm and research design (Ellis & Barkhuizen, 2005). In the current research, the design was
so strongly linked to my beliefs regarding SLA theory that I arrived at my research paradigm by
default.

4.2: Research Designs in Three Classroom Studies

This section explores the research design used in three prior L2 writing studies. The first two
were conducted in secondary school FFL classrooms with similar contexts to the current study.
The third was conducted in advanced level university ESL classes and was framed by the
sociocultural-interactionist perspective on SLA that has been selected for the current study.

4.2.1: Study One – Benevento and Storch, 2011 (Australia)

Although not specifically stated, it seems that Benevento and Storch (2011) conducted their
study from an interactionist stance on SLA. They asked: Does L2 writing among secondary school
French learners develop or improve over a period of six months? What aspects of writing, if any,
develop or improve? Their purpose was to challenge the assumption that learners develop
language and writing skills in response to teacher feedback, instruction and practice. The
research context was naturalistic and the data clinically elicited. The participants were fifteen
students in their final year of secondary school who had been studying French for five years, four
lessons a week. For the Victorian Certificate of Education (VCE) high-stakes examination for
which these students were preparing, students are required to submit school-assessed
coursework and homework as well as sit external examinations. The coursework used for the
research data involved writing a variety of text-types in French in response to real-world
questions.

While the data were being collected, the teacher included grammar as well as communicative
skills in the programme because he believed that accuracy in increasingly complex structures
enabled students to gain higher marks. In preparation for in-class writing assessments, the
teacher asked students to brainstorm ideas, create useful vocabulary and structure lists, look at

previously written models and practise writing different text-types. He also provided explicit individual feedback on errors found in drafts and whole-class instruction on common errors. The summative evidence was gathered under test conditions on three separate occasions over six months as required by the examination authority. On each occasion students had between fifty and sixty minutes to write between 200 and 300 words and were permitted to use a dictionary.

The teacher graded the texts according to examination specifications, taking into account vocabulary and structures, adherence to text-type, originality of ideas, and overall cohesiveness. The researchers measured the language in the scripts for complexity using T-units (main clauses with their dependent clauses) and tallied connectors and lack of accuracy with error codes. Independent raters established reliability scores of 100% for complexity and 97% for accuracy. In the process it became clear that many students had used error-free chunks of language learned in class. Therefore, to ascertain the error rates without these chunks, the chunks were tallied also.

Benevento and Storch (2011) found that the students' writing improved over time but with variations. Syntactic complexity increased between Tasks 1 and 2 but decreased between Tasks 2 and 3 while the use of connectors increased in Task 3 only; the organisation of arguments, the structure of texts and the richness of vocabulary improved over the three tasks; in their third text, students did not use many formulaic chunks; from Task 1 to Task 3, accuracy did not increase; and errors in gender, articles, prepositions and word choice remained consistent.

The researchers considered the fact that their data came from scripts written by students for authentic classroom assessment to be a strength as well as a weakness. While it gave the research ecological validity, the Victorian qualification required students to write different text-types of slightly different lengths under slightly different time allowances which did not allow for the direct comparisons that pre-and post-tests usually require. Syntactic accuracy could therefore have been affected. In addition, the researchers questioned the wisdom of giving feedback on all errors when targeted feedback is more effective (Bitchener, 2008).

4.2.2: Study Two – Graham and Macaro, 2007 (Britain)

Study Two was an intervention study conducted and reported by Graham (2004) and Graham and Macaro (2007), seemingly from an interactionist stance on SLA. The researchers used mixed methods to collect data in a communicative and Assessment *for* Learning context which featured process writing, peer and teacher feedback, strategy training and reflection journals. This

extensive British study researched the impact of strategy training on the FFL listening and writing achievement of 257 sixteen to seventeen year old Year 12 students in ten schools over seven months. The ultimate goal was to provide direction for British education policy by investigating whether self-efficacy improved when learners were trained to notice the relationship between their strategy use and their achievement. Quantitative data were collected from pre- and post-intervention tests and qualitative data from reflection journals and teacher feedback on strategy use (Graham, 2004; Graham & Macaro, 2007).

Three constructs were explicitly defined in the research report. The first, self-efficacy (an aspect of agency), comes from Bandura's (1977) social cognitive theory and describes the belief that learners hold concerning their perceived ability to succeed, which can be different from their actual ability to succeed. According to Bandura, those with high self-efficacy who believe they are able to perform well in future tasks are more likely to do so; those with low self-efficacy who believe future tasks to be difficult are more likely to avoid the tasks or fail in performing them. The second construct was language learner strategies, the conscious linguistic and metalinguistic thought processes evidenced in the learners' writing when they took steps to solve language problems, respond to feedback, and make improvements in their work. The third construct was instrumentality, the link between strategy use, attributions (or reasons learners gave for their success or failure in certain tasks) and whether they chose to continue learning the language or not. Many research studies (Cohen, 2011; Cohen & Macaro, 2007; Dörnyei, 2001; Graham, 2004) have shown that good language learners use a wide range of metacognitive strategies to aid in their learning and that strategy use is directly linked with motivation to learn, in particular, the motivation to persevere in the face of difficulties and to continue learning languages into the senior school. If learners believe that their performance can be explained by matters within their control, such as the amount of effort they put in or the type of strategies they use, they are more likely to attempt the task again. If they feel that their success or behaviour is beyond their control then they are less likely to try again (Graham, 2004; Williams & Burden, 1997).

The research involved three groups. One group of students from four schools (50 pre-test and 46 post-test) completed a range of listening comprehension and writing assessments; they formed the control group. A second group of students from five different schools (54 pre-test and 40 post-test) completed the assessments and received detailed strategy instruction. A third group of students from five schools (35 pre-test and 32 post-test) completed the assessments, received

strategy training, and, in addition, kept monthly entries in reflective diaries on their strategy use. Groups Two and Three took part in the same strategy training activities although Group Three was given an additional consciousness-raising activity at the start of the project. Only Group Three students reflected on what went well and why, what did not go well and why, what they felt about their strategy use and their writing, and what they thought they needed to do next. In addition, they were asked to give feedback on the content of a peer's writing, to complete a practice task based on their error feedback, and to read the teachers' comments on strategy use and linguistic accuracy. The comments were designed to draw attention to the links between strategy use and successful writing performance.

Students in all three groups took part in pre- and post-intervention listening and writing tests then answered a self-efficacy questionnaire which explored how certain they were of achieving well. In order to understand students' attributions, at the end of the project they were asked to complete another questionnaire in which they specified their successes and failures and gave reasons for them. They were also asked to state how much they thought they had improved and the role they felt that the strategy training had played in that improvement.

The research found that all students made some progress in self-efficacy over the six months, but the third group that received the most scaffolding (strategy instruction and feedback on diary reflections) made the most progress followed by the second group and then the first. However, the actual achievement scores of the second and third groups were similar. Forty-seven percent of students felt that the strategy instruction had helped their performance.

Even though the study focussed on self-efficacy, the researchers were able to advise teachers regarding strategy instruction related to the formulation stage of process writing which the students had found difficult. The researchers described formulation as the moment when the writer converts an idea into language, the moment when students have many choices to make, for example, how to overcome difficulties, and which language to use when thinking about difficulties. They found that very little research attention has been paid to the formulation stage of L2 writing and "[e]ven less research has been conducted into how learners might be helped to adopt more effective strategies for the formulation stage, and certainly none has taken place in contexts similar to that of post-16 in England" (Graham & Macaro, 2007, p. 154).

As a result of their findings, Graham and Macaro suggested that teachers should begin writing instruction programmes by asking students to write on topics not yet studied and with access to only a dictionary. By encouraging students to take risks, they are likely to discover their students' language problems and the strategies they use to overcome them. They suggested treating each stage in the writing process separately by encouraging students to reflect on how they plan, how they go from an idea to a first draft, how often they look up words, how they monitor for mistakes (new and residual), and how they deal with feedback. Finally, they advised teachers to ask students to evaluate how this process helps them write better. Graham and Macaro concluded that there was something to be gained from strategy instruction that begins early in the year despite the time it takes, but they hesitated to turn their suggestions into recommendations for national policy.

4.2.3: Study Three – Wigglesworth and Storch, 2012 (Australia)

This third study was a mixed method intervention conducted by Wigglesworth and Storch (2012) with 36 pairs of advanced ESL university student volunteers who were paid to take part. They researched four matters: the extent to which three types of feedback allowed the students to improve their writing; which of the three types was the most effective in terms of improved accuracy; the extent to which the students engaged with the feedback; and any other factors that might impact upon their engagement. They defined the construct of engagement as involvement in language related episodes, but acknowledged the difficulty of ascertaining levels of engagement through speech alone as one partner might still have been silently engaged in a language related episode while the other may have already moved on to the next. They defined language learning as the number of learning opportunities which occurred and were taken up as the pairs discussed feedback on their texts as well as the improvements to accuracy and complexity in the texts that the individuals produced.

Wigglesworth and Storch (2012) chose to gather evidence from pair writing because, from a sociocultural perspective, collaboration can activate learning as learners pool their linguistic resources and scaffold each other's progress. They assumed that collaboration would produce better written output than either individual might be able to produce alone; that discussing feedback with a partner would alert writers to aspects of language not previously noticed and engage them in discussion about language (language related episodes) which would advance their language knowledge; that collaborative dialogue initiated by feedback would deepen

engagement and enhance learning; and that mixed results from previous feedback research had emerged because students react differently to different types of feedback.

The researchers gathered data in three sessions over a four week period. During the first session, the self-selected pairs collaboratively composed a written report. Five days later they received one of three types of feedback on their writing, either no feedback, reformulation (explicit error correction), or editing (indirect codes that indicated an error but did not correct the error). As they wrote their first draft in the first session and edited their writing based on the feedback in the second session, the participants' conversations were recorded and then transcribed. Four weeks later, individuals were asked to rewrite the report under test conditions using just the first draft and without resources or access to any previous feedback.

When analysing the scripts, Wigglesworth and Storch (2012) found that neither forms of feedback, editing or reformulation, led to improvements in complexity but that both forms of feedback led to substantial improvements in accuracy, with reformulation the more effective of the two conditions. When analysing the levels of engagement in language related episodes, they found that engagement with both types of feedback was high but that it was higher for the editing group than the reformulation group. However, this high level of engagement in the editing group did not lead to more accuracy. They suggested that, because there was evidence in the data that several pairs in both groups used memorisation as a strategy when responding to feedback in preparation for the rewriting test, it would seem logical that those memorising the explicit reformulation feedback would produce more accurate language than those who memorised language they deduced from indirect feedback. They concluded that learners play an important role in their own learning and that inviting learners to write in pairs increases their writing development because it gives them "more, and more in depth, learning opportunities in terms of their language use and their insights into language than they might otherwise encounter" (p. 92).

4.2.4: How the Three Classroom Studies Informed My Research Design

Several insights were gained from the above three classroom studies for the present study, including the format of the writing units and tasks, the measures used, and the interpretation of the findings. In the research conducted by Benevento and Storch (2011) the context was very similar to the context of this present study; it was high-stakes, secondary school portfolio assessment that included tests as well as homework, and communicative tasks as well as

grammar instruction. Firstly, the researchers' misgivings about giving feedback on every error influenced the current research design. Although indirect coded feedback was given on all errors, participants were asked to tally their error types to find their two or three most frequent and to focus on just these. Secondly, Benevento and Storch chose to measure language development using a combination of interactionist text-analysis tools to measure complexity and accuracy. Instead of a measure of fluency, they used a more subjective holistic rubric required by the state assessment to measure overall communicative quality. Although it was my original intention to use this combination of measures, queries at a conference where I presented my unfinished work inspired a change of mind. I therefore added a suitable fluency measure to complete the complexity, accuracy, fluency trio, and complemented these with an analytic measure of L2 written proficiency, using the holistic national rubric to place participants into pre-test groups only. This provided a second opportunity after the pilot study (see Chapter Three) to fill the research gap reported by Wolfe-Quintero, et al. (1998) by using all three CAF measures with the same group of participants. Thirdly, the researchers found that while complexity and overall quality increased in the students' L2 writing over time, accuracy did not. Based on their opinion that this resulted from comparing different text-types, I was careful in my research to compare increase within the same text-type before comparing across text-types (Benevento & Storch, 2011; Kuiken & Vedder, 2008; Nunan & Bailey, 2009; Schoonen, et al., 2009).

The second study, conducted by Graham and Macaro (2007), measured the gains in self-efficacy that resulted from different amounts and types of strategy training, including via the use of a metacognitive reflection diary. Three factors influenced the present research. Firstly, the students reflected in diaries on how they planned, how they went from an idea to the first draft, how often they looked words up, how they monitored for mistakes (new and residual), and how they dealt with feedback. The content of these reflection diaries was an important consideration for the reflective blogs used in the present research. Secondly, the researchers debated the value of strategy training because it took time away from instruction time, which was the very issue that motivated the present research. They concluded that the gains in self-efficacy alone were well worth the time taken. Thirdly, they found that the large number of teachers who conducted the research activities threatened data reliability. In the present research design, therefore, the researcher rather than the four classroom teachers conducted all research activities in the five classrooms.

The third study conducted by Wigglesworth and Storch (2012) was of interest to the present study for four reasons. Firstly, the researchers used a combination of sociocultural and cognitive-interactionist tools to measure L2 learning, which they defined as the level of engagement in language related episodes as well the improvement in complexity and accuracy between drafts. Secondly, it was of interest because the researchers felt that collaborating in pairs produced better written output than either individual could produce alone; they discussed feedback with a partner which alerted them to aspects of language they might otherwise not have noticed and they engaged in discussion about language which advanced their language knowledge, deepened engagement, and, in turn, enhanced learning. I wondered whether this collaboration might be particularly effective for learners at an intermental stage of language learning and whether an individual learner reflecting on feedback might have similar results if they were at the intramental stages of language learning (Aljaafreh & Lantolf, 1994). Thirdly, the research was of interest because of the ongoing feedback debate. The researchers found that explicit feedback resulted in more accuracy gains than indirect feedback and they suggested why this might be the case. They found that several pairs of writers engaged in the strategy of memorisation as they received feedback on their first draft and prepared to rewrite their second draft under test conditions. It was a reminder that the use of tests tends to invite memorisation (Sachs & Polio, 2007). Finally, the research was interesting because both feedback conditions resulted in an increase in accuracy but not in complexity, which was opposite to the findings in Benevento and Storch (2011) who researched over a longer period of time. It is possible, that, in portfolio writing contexts, when measuring gains between the drafts of one task written within a short time period, increases in complexity are not likely. An increase in complexity seems more likely between different tasks and written over a longer time period.

4.3: Research Design

In Chapter One, I detailed my beliefs regarding SLA theory and discovered, as a result, my position on the research paradigm continuum (see Section 4.1 above). Now, following the literature review conducted in Chapter Two, the pilot study in Chapter Three, and the insights from the three classroom studies in Section 4.2 above, I detail the mixed method, counter-balanced, comparative design considered to be the most appropriate for answering my research questions. In the next three sections I follow the six steps recommended by Norris and Ortega (2005) by: defining the constructs to be measured and identifying the behaviours to be observed;

specifying the tasks that will elicit such behaviours and then eliciting, recording and observing the behaviours while controlling or accounting for variables; scoring the observing behaviours in a way that connects them to interpretations, and summarising and analysing the data according to probable categories.

4.3.1: Construct Definitions and the Behaviours that Represent Them

A construct is a psychological quality that is not directly observable. We assume these constructs exist because we observe the results of their existence in the behaviour of our research participants (Norris & Ortega, 2005). To establish construct validity, the constructs important to this research are now defined so other researchers might recognise them in my work and evaluate whether the observable behaviour is truly evidence of the unobservable construct (Nunan, 1992). Constructs that are central to the current research include *L1 reflection, L2 learning, language learner strategies, self-regulation, and motivation.*

L1 reflection refers to the immediate, retrospective, guided report written by the participants in English and in shared Google Docs. In these reflective blogs, the participants pondered the processes and strategies they used when writing in FFL, they recorded their reactions to feedback on their writing, and they responded by setting goals. Although incomplete, the content of the blogs was assumed to represent the participants' cognitive processing. In this L2 process writing context, it equated to auto-feedback (Ellis, 2005), self-scaffolding, private speech and languaging (Knouzi, et al., 2010), and dialogic engagement between students and their teachers (Feuerstein & Feuerstein, 1999; Slavkov, 2015). According to previous scholarship, such reflective practices provide opportunity for: students to become independent learners (Ellis, 2005) and self-regulated learners (Schunk & Usher, 2013; Zimmerman, 1990); heightened awareness that results in engagement, empowerment and enjoyment (Houston & Turner, 2007; Langer, 2000); attention to feedback (Polio, 2012); and the type of deep metalinguistic processing that determines the significance of L2 outcomes (Kuiken & Vedder, 2008; Manchón, 2011a & 2011b).

L2 learning is defined as the single-event language changes that participants made to their second drafts as a result of feedback and reflection on their first drafts. *L2 learning* compares with L2 *development* which is defined as language changes that result from multiple mediations over time and which require increasingly less assistance (Aljaafreh & Lantolf, 1994); and it compares with L2 *acquisition* which is defined as the learner's consistent, coherent, accurate and fluent

language produced without needing support, and which is fully internalised and self-regulated (Norris & Ortega, 2005). Although, in the research, L2 learning gains were measured, snapshots of the participants' L2 development and L2 acquisition as it related to NZC Level 6 verb knowledge were also able to be recorded using an adapted version of Aljaafreh and Lantolf's (1994) Regulatory Scale. From these snapshots, it was assumed that L2 acquisition had occurred when the participants were able to produce certain verb types accurately and consistently or, when mistakes were quickly self-corrected. It was assumed that participants were in the process of developing verb knowledge and approaching the self-regulation end of the scale when they were able to locate errors in their first drafts as a result of feedback and make appropriate changes in their second drafts without further support. When learners were able to locate their errors and make correct changes in response to feedback and with the aid of resources, it was assumed that they were nearer the middle of the self-regulation scale; when they located the errors but avoided making changes or edited them wrongly as a result of the feedback, they were assumed to be nearer the other-regulation end of the scale; and when they failed to locate errors despite feedback, it was assumed that the language feature was outside of their ZPD.

The third construct for definition is *language learner strategies* which are defined as the "specific actions taken by the learner to make learning easier, faster, more enjoyable, more self-directed, more effective and more transferable to new situations" (Oxford, 2001, p. 166), and the "[a]ctivities consciously chosen by learners for the purpose of regulating their own language learning" (Griffiths, 2008, p. 87). Strategies are closely linked with *self-regulation* because they are one of the means that language learners use to control and actively participate in their own learning (Wendon, 1999). Teng and Zhang (2016) established the connections between self-regulated learning, writing strategies and written performance. In the context of the current research they are the actions and approaches reported in the learners' L1 reflective blogs and evidenced in their L2 writing as they implemented plans, overcame language problems, responded to feedback, evaluated and improved their writing, set goals for improving future writing and indicated their journey towards independent learning.

The fourth construct is *self-regulated learning* itself. The metacognitive, cognitive, affective and social-interactive strategies that the participants reported using as they planned, composed and reviewed their L2 writing, and the motivations, feelings of self-efficacy and goals they set to improve their L2 writing are all components of self-regulation because they "underlie self-initiated

and self-sustained efforts to acquire competencies" and "acquiring mastery of important academic and non-academic skills has been termed self-regulated learning" (Zimmerman, 2013, p. ix).

The fifth construct is *motivation*, the intrinsic and extrinsic reasons why students maintain effort and persevere when learning a language and why they choose to continue learning from one school year to the next. Ushioda (2009) suggested that the best way to capture motivation is by examining classroom discourse rather than by questionnaire. In the current research, therefore, this discourse is the reflective blogs. The participants' motivation is encountered in the goals they set for improving their writing and, ultimately, in the choices they made to continue or not continue learning French at the end of the school year.

4.3.2: Tasks to Elicit These Behaviours

The L2 data for the research were elicited by two communicative writing tasks (see Appendices C and D) and a counterbalanced research design. These produced evidence that was able to also be used for the participants' NCEA assessment portfolios. The dual purpose minimised intrusion into the learning programmes in the four participant schools and ensured a high level of student engagement with the research. According to Ellis and Barkhuizen (2005), the data had ecological validity as the evidence was produced for an authentic national assessment by authentic students in an authentic school context; the data also had psychological validity because the language was produced in a communicative context where meaning and message were primary concerns, and it was the sort of language that would be produced in real life.

The first task entailed writing a personal profile and seeking an e-pal, which is an exercise suitable in the first term of Year 11 for reviewing the previous year's language and showcasing early Year 11 language. The second task entailed writing an online blog entry detailing a day's events while on holiday in a special place. It was suitable for showcasing language typically learned in the first half of Year 11 and required by the writing standard (see Appendix A), that is, "language related to basic personal information and past, present, and/or future experiences, in order to express personal information, ideas and opinions in culturally appropriate written French" (NZQA, 2015, online).

Participants responded to these tasks in two four-week process writing units held three months apart (see Table 12) and produced four drafts each, 284 drafts in total. Collecting data from two writing units enabled a counterbalanced research design that ensured no participant was

advantaged or disadvantaged by the research. In the writing unit that took place at the end of

Term 1, Schools A and B followed a sequence of writing test, feedback-with-reflection, redraft,

while Schools C and D followed a sequence of writing test, feedback, redraft. Three months later,

at the end of Term 2, the sequence was reversed. Schools C and D received the reflection

intervention while Schools A and B did not.

Table 12: Process Writing and Intervention Sequence

Term 1	Schools A and B	Schools C and D
Week 1	• Teacher distributed task instructions and assessment schedule one week before Draft 1 writing test • Students planned	• Teacher distributed task instructions and assessment schedule one week before Draft 1 writing test • Students planned
Week 2	• Students wrote timed (30 minute) first drafts with no resources • **Students responded to the prompts in the first part of the guided reflection**	• Students wrote timed (30 minute) first drafts with no resources
Week 3	• Teacher provided coded feedback in the margins of the first drafts and feed forward at the ends of the drafts • Students tallied their error codes to discover the two or three most frequent error types • Students had ten minutes of silence to edit based on teacher feedback • Teacher instructed class (Focus-on-Form) on common errors • Students had ten more minutes to edit silently based on teaching • Students selected practice activities on their three most frequent errors for homework to aid further editing • **Students completed the guided reflection** • Students took their annotated first draft home and used it as a prompt to rewrite their second draft	• Teacher provided coded feedback in the margins of the first drafts and feed forward at the ends of the drafts • Students tallied their error codes to discover the two or three most frequent error types • Students had ten minutes of silence to edit based on teacher feedback • Teacher instructed class (Focus-on-Form) on common errors • Students had ten more minutes to edit silently based on teaching • Students selected practice activities on their three most frequent errors for homework to aid further editing • Students took their annotated first draft home and used it as a prompt to rewrite their second draft
Week 4	• Seven days later students submitted second draft	• Seven days later students submitted second draft

Each writing unit was considered an intact event. The first drafts in each unit acted as pre-tests

for the intervention while the second drafts in the same unit acted as post-tests. In both process

writing units, participants were given the task one week in advance of the writing test (see Table

12). In the test, participants had thirty minutes to write their first draft without access to resources.

A week later, in a dedicated lesson, first drafts were returned with feedback in the form of indirect

codes in the margin of the line in which errors occurred (Ferris, 2006; NZQA, 2006) and feed

forward comments at the bottom of scripts (NZQA, 2006) which included an NCEA grade

indicator. Along with their first draft and feedback, participants received a copy of the coding

schedule (see Appendix F) which contained definitions of the codes, question prompts that might

help locate an error, an example of a strategy that might help in repairing the error, and/or a link

to relevant practice exercises related to the error type.

To begin the review process, participants were invited to tally their error codes and find their two

or three most frequent error types. This allowed them to focus attention on just a few errors

(Bitchener, 2008) and to be motivated by involvement in the process (I. Lee, 2007). The tally was

followed by ten minutes under test conditions and without resources for the participants to begin

editing their drafts in response to the feedback. Any changes made to the drafts during this first

edit period were marked with a straight line (see Figure 1).

Figure 1: Script with initial edits from indirect codes and teaching on common errors

This first edit was followed by fifteen minutes of teaching on the five most common errors per

class which was designed to benefit all members of the class (Colby-Kelly & Turner, 2007; Fotos

& Hinkel, 2007; Harley, 1989), and then a second opportunity to edit. On this second edit,

changes were marked with a wavy line (see Figure 1). Ellis and Shintani (2014) equated editing in

response to feedback as 'uptake', an important step in the language learning process. Edited

drafts were then collected, copied and returned to participants at the end of the lesson. At that

point, control groups were invited to take their copy home for further rewriting with access to any

resources they chose (as the NCEA assessment standard specified), while intervention groups

were given an extra ten minutes to ponder the review process and respond to the template

prompts (see Appendix E). If participants were unable to repair their drafts based on the feedback

alone, the error code schedule (see Appendix F) directed them to strategies and practice exercises to help them develop an understanding of the language rules associated with their errors (Ellis & Shintani, 2014). One week later, participants submitted their second draft.

The L1 reflection data were elicited via the reflection intervention, which gave the participants the time and occasion to reflect using the guiding template (see Appendix E). As explained above, the reflection occurred in two phases: the participants responded to the template's first two sets of prompts immediately after writing the first draft, and in the same one hour class. They responded to the second set of template prompts immediately after responding to feedback in the one hour feedback lesson. In the first phase (which occurred after the writing test and while their memory remained fresh) participants were prompted to reflect on how well they had planned for their first draft and on the strategies they used to overcome any difficulties they encountered during the test, and to set goals to improve subsequent planning and composing processes. In the second phase (at the end of the feedback lesson), participants were prompted to reflect on their reactions to the feedback they had received and to set goals to improve subsequent reviewing processes.

4.3.3: Template to Elicit the L1 Reflection Data

When developing the L1 reflection template, three factors were considered: firstly, the maturity of the participants and their ability to analyse and think abstractly; secondly, the topics raised through the literature review that have the most potential to promote learning (see Sections 2.4.2, 2.4.3 and 2.4.4); and thirdly, the stages of process writing and a consideration of the best places to pause and reflect in the process writing sequence.

Firstly, as explained by Porto (2007) and Anderson (2008), adolescents have a developing ability for abstract reflection but need guidance that indicates to them what is important to consider for learning. The written reflection was, therefore, viewed as a form of inner speech (see Section 1.3.3), and making time to reflect using a template as scaffolding typical of the middle of three stages of microgenesis. Aljaafreh and Lantolf (1994) referred to this middle stage as partially self-regulated or partially intramental because the learner no longer needs the explicit guidance of the first stage but is not yet able to reflect implicitly and fluently without teacher support (the third stage). Grouped questions or prompts were therefore used rather than a first stage question and answer format or a third stage open-ended format.

Secondly, when considering the topics for reflection that are most likely to promote learning (see Section 2.4.2), previous scholarship suggested learning habits (Belgrad, 2013; Grift, 2007), goal setting (Feuerstein & Feuerstein, 1999; I. Lee, 2007), metalinguistic reflection on feedback (Bitchener, 2008; Manchón, 2011a), and learning and communication strategies (Anderson, 2002; Chamot, et al., 2005; Oxford, 2011; Zhang, 2013). Such topics were also modelled in research conducted by Graham and Macaro (2007) and Oxford (2011), and suggested for classroom use by Williams and Burden (1997). From my own classroom practice, I was aware that, while students of this age are often happy to receive feedback, they do not always act upon it to improve their work. I wanted to use the reflection in such a way that they would be motivated to write an improved second draft, and, according to Assessment *for* Learning principles, learn language in the process.

Thirdly, the three dynamic stages of process writing (planning, composing, and reviewing), the cognitive load of reflecting when writing for assessment, and the value of reflecting while memories remained fresh were considered. As the planning and rewriting stages in the process writing sequence were able to occur at home, the first draft writing and the feedback lesson occurred at school. The reflection template was consequently divided into two parts, one part for each of those lessons. Self-regulation and agency were encouraged by involving students in the feedback process as much as possible (I. Lee, 2007). The first half of the reflection template (see Appendix E) was recorded immediately after writing the first draft. By reflecting immediately after L2 writing rather than during writing, the cognitive load of both activities was managed yet memories remained fresh. During this first part, the template prompted students to reflect on how well they had prepared (e.g. length of time, strategies used, the process they went through) and how happy they were with their preparation in light of the writing they had just completed in class, and what they wanted to remember when planning to write in the future. Then they were prompted to reflect on any difficulties they might have experienced while writing, on the strategies they used to overcome those difficulties, on what had gone well, and on what they wanted to remember for next time. The second half of the reflection was recorded seven days later and immediately after receiving feedback on the first draft. Students were prompted to tally their most frequent error types as indicated by the coded feedback. Before rewriting, and to strengthen their understanding of the grammar rules related to their most frequent types of error (Ellis & Shintani, 2014), students were invited to choose remedial practice exercises and to reflect regarding them. They were also asked to reflect on the effectiveness of the four types

of feedback offered (indirect codes, teaching on common errors, teacher's comments, and NCEA grade indicator). Finally, students were invited to reflect on anything further regarding the processes and strategies they had used or intended to use when writing the first draft and subsequent drafts. This provided a less supported opportunity for those students who might be closer to the fully self-regulated stage to reflect on matters important to them.

4.3.4: Collecting the Data while Controlling and Accounting for Variables

The data collection methods detailed above required careful consideration because of inherent difficulties when conducting intervention research in L2 classrooms. It was important to establish data validity by carefully linking the intervention with any research phenomena, that is, the "cause and effect which is particularly problematic in open and dynamic systems like schools" (Lai, 2013). It was also important to establish data reliability by showing the consistency of the data collection, analysis and interpretation and the extent to which the research could be replicated (Nunan, 1992).

Key threats to reliability and validity in classroom contexts typically include poor research design, using learner language as evidence, and practitioners researching students (Norris & Ortega, 2005). To control for these threats, Norris and Ortega recommended matching a straightforward and unambiguous research question with data and data collection methods, randomly assigning control and intervention groups, accounting for halo effect and teacher beliefs, paying attention to task design, using triangulation and pilot studies, and employing objective raters to moderate work independently. These recommendations were followed as much as was possible in the context and are now discussed in turn.

In initial meetings with supervisors, the research questions were carefully considered to ensure they remained straightforward, specific and unambiguous. Although they were adjusted on several occasions, the essence remained the same: Does L1 written reflection increase L2 learning when adolescents write for portfolio assessment? Matching the data with the question was the next consideration. The research design invited participants in intervention groups to reflect on their L2 writing while participants in control groups did not. The changes that occurred between drafts, following the definitions in Sections 1.3.3 and 4.3.1, were considered to be L2 learning. Any difference in learning gains between the intervention and control groups was considered to be a possible effect of the reflection intervention.

Matching the data with the data collection methods was not quite so straightforward. The NZQA (2015) writing assessment specifications allow students to access resources as they write and to improve their scripts over time (see Appendix A). These assessment exigencies necessarily affected the data collection process. While the first draft of each task was written under test conditions without resources and acted as a pre-test, the second draft, the post-test, was written at home with resources which fulfilled the requirements of the writing standard. It should be noted that students were given the entire task a week in advance of the test so they were able to plan for their first draft with access to resources just as they had access to resources for their second draft. Consequently, very few participants made any changes to their second drafts other than word-level edits. It appeared that the different conditions under which the first and second drafts were written had minimal effect on the data collected.

To further ensure data validity, the procedures for the data collection were carefully considered both for the research and to ensure minimal interruption to the Year 11 learning programmes in this, the participants' first national examination year. To mitigate the risk of researching in four different schools and in five different classrooms taught by four different teachers, timing was kept tight. Participants were given the same task instructions and the same number of days (seven) to prepare for the first draft test, they wrote the first draft in thirty minutes under test conditions, and feedback on the first draft was delivered in a dedicated lesson exactly seven days later. Then all students had a further seven days to improve their writing in their own time before submitting their second draft. Exactly the same process occurred in Term 2 as in Term 1. Of course, during those writing units, lessons not used for research purposes were taught by classroom teachers and contributed to curriculum coverage.

In addition to the stringent timings for the data collection, the researcher rather than the classroom teachers supervised all the writing tests, coded and gave feedback on all the L2 writing drafts and conducted all feedback lessons. This controlled for halo effect, the bias in teacher judgment that results from an unconscious influence of unrelated matters, such as students' personality traits or the learning history between the student and teacher. The teachers' role became purely administrative; they gave initial permission for their students to take part in the research and they undertook important administrative tasks such as the timetabling and rooming of research lessons, access to computers, the distribution of task information, photocopying, and the collation and distribution of evidence.

One strategy often used to mitigate risk in classroom research is the random assignment of intervention and control groups (Norris & Ortega, 2005; Nunan, 1992). In the current research this was not possible because students were already placed in intact classes by their schools, which, in two of the four schools researched, were the only Year 11 French classes. Instead, the assignment of intervention and 'control' groups occurred according to geographical location and convenience. Two schools in one sector of the city (one decile 4 and the other decile 7[10]) formed the Term 1 intervention group, and two in another sector of the same city (one decile 6 and the other decile 10), formed the 'control' group. Because of the lack of randomisation, however, this study must be referred to as a comparative study rather than an experimental study and the 'control' group is more appropriately referred to as a comparison group. None-the-less, the pre- and post-tests typical of experimental designs remained important features in this counter-balanced comparative study (Martella, et al., 2013; Nunan, 1992). Despite lack of randomisation, internal validity (the claim that the outcomes are a result of the intervention) was strengthened by the shared characteristics of the participant schools and students; they were Year 11, FFL students in state, coeducational, secondary schools, who had opted into their non-streamed classes by choice, who were taught using communicative language teaching methodologies, and who wrote for the NCEA Level 1 writing standard using NZC Levels 5 and 6 language.

A further threat to reliability and validity that needed to be considered was using variable and inconsistent learner language as evidence. Different task designs, especially the choice of genre or text-type, have been shown to influence the micro-and macro-linguistic characteristics of the language produced (Benevento & Storch, 2011). This risk was controlled by asking the four participant teachers to collaborate in the writing of the tasks so that they suited the learning programmes in all four schools and so that participants wrote the same text-types (i.e. an online profile and a travel blog) in the same writing unit. Using learner language was also controlled by firstly comparing the learning gains between intervention and comparison groups when completing the same task (intergroup) before comparing both comparison groups with both intervention groups who completed different tasks (intragroup).

---

[10] Deciles indicate the socio-economic status of a school's student body and are calculated every five years based on census data. For example, the 10% of schools that have most students from the lowest socio-economic communities will be decile 1, while the 10% of schools that have most students from the highest socio-economic communities will be decile 10. The decile system is for funding purposes only and ensures that those who have the biggest barriers to overcome receive the most financial support. It is important to note that a school's decile rating does not indicate the quality of education delivered (MOE, 2016, 29 November).

As well as ensuring data is valid and reliable, East (2008) stated that researchers in CLT and assessment contexts need to ensure data collected is authentic and fair; it needs to relate to how language is used in the real world; and it needs to "provide adequate opportunity for all test-takers to demonstrate their best performance" (p. 26). Authenticity was therefore ensured by inviting participating teachers to contribute collaboratively to the writing of the real-world communicative assessment tasks. Fairness was ensured, firstly, by the counterbalanced research design, the Term 1 intervention group becoming the Term 2 comparison group and the Term 1 comparison group becoming the Term 2 intervention group, and secondly, by gaining ethics approval from the university.

Ethics approval is important when researching with human participants. Permission for access to the students' writing and reflections was firstly gained from the four schools' Principals and Boards of Trustees. Although there is no law in New Zealand that requires informed consent (Loveridge, 2010), the University of Auckland Human Participants Ethics Committee insisted that consent should also be attained from parents and guardians of students under the age of sixteen, and that assent be attained from students because they were of an age (most were fifteen) to understand the project. Informed consent information included the specific details of the project: the storage, retention and destruction of data, the publication of results, a guarantee of anonymity and confidentiality, and the right of withdrawal at any time. The information was written in language that was clear, appropriate and straightforward. This guaranteed no deception of any kind. The information also included a promise of no harm, so all students were offered the opportunity of the reflection intervention, two classes in Term 1 and three classes in Term 2 (see Appendix M).

Norris and Ortega (2005) also recommended controlling risks in classroom contexts by conducting pilot studies, triangulating the collection and measurement of data, and employing objective raters to independently moderate work. The pilot study reported in the previous chapter was conducted to find the most suitable quantitative measures for the context (see Figure 2). The data collection was triangulated by supporting L2 data with L1 written reports as well as by measuring L1 and L2 data using qualitative and quantitative measuring tools. The next section details how the L1 and L2 data were scored, how the results were summarised and analysed, and how inter-rater reliability scores were obtained.

4.3.5: Scoring, Summarising and Analysing Observations

Before the scoring began, the NCEA writing assessment schedule (see Appendix B) was employed to grade the first drafts of both writing tasks in order to give feedforward to the students and to provide pre-test data for the research. Using the NCEA rubric, the participants were placed into three achievement groups (Achieved, Merit and Excellence) to facilitate a more fine-grained analysis (Budoff, 1968; Sasaki, 2000). (This meant that, although similar, there were slightly different pre-test group sizes for each task.)

However, because the NCEA rubric is a relatively blunt measuring instrument (there are only four grades to award, including Not Achieved), and because 56% of the participants gained the same grade in their first draft of Task 1 in Term 1 as they gained for their final end-of-year portfolio, other measures capable of discerning discrete changes between drafts were required. A quadrant of measures was consequently developed (see Figure 2). The quadrant balanced measuring tools according to those that located learning in the L2 writing (upper quarters of the quadrant) and those that revealed insights in the L1 reflective blogs (lower quarters of the quadrant). The quadrant also balanced measures that quantified (left-hand quarters) with measures that required training and expert judgment (right-hand quarters).

To measure learning in the L2 writing scripts, objective tools included CAF analysis and error analysis, while those tools requiring specialised training and judgment included East's (2008) Scoring Rubric and Aljaafreh and Lantolf's (1994) Regulatory Scale. The qualitative coding of the reflective blogs was guided by the lists of strategies compiled by Zimmerman (1990), Graham and Macaro (2007), and Oxford (2011), while Biggs and Collis' (1982) Structure of Observed Learning Outcomes (SOLO) Taxonomy provided a five level rubric for measuring depth of thinking. These tools which required expert judgment were supported by the digital tools, Word (Version 14.0.7173.5000, Microsoft Home & Student, 2010), OneNote (Version 16.0.7127.1026, Microsoft Home and Student, 2013), and Wordsift 2 (www.wordsift.org). By using this variety of measures, the reliability of the data was strengthened and different facets of language learning were investigated.

Following the pilot study reported in Chapter Three, L2 writing samples were scored quantitatively using situated CAF analysis measures: the ratio of dependent clauses per clause measured complexity, the ratio of error-free clauses per clause measured accuracy, and the number of

error-free clauses per composition measured fluency. These were found to be suitable for

measuring data of a low level and of the same level.

Figure 2: Quadrant of Measures

Locating L2 Learning in the French Writing Samples

Text-analysis using measures that quantify:

1. CAF Analysis
2. Error Analysis

Text-analysis using measures that require judgment:

1. NCEA Rubric
2. East's Rubric
3. Aljaafreh and Lantolf's Scale

Quantitative/Objective

Qualitative/Subjective

Coding using digital tools that quantify and collate:

1. Wientjes & Hakuta's Wordsift
2. Microsoft OneNote
3. Microsoft Word

Coding using lists and rubrics that require judgment:

1. Zimmerman's Strategies
2. Graham and Macaro's Strategies
3. Oxford's Strategies
4. Biggs & Collis' SOLO Taxonomy

Revealing Insights in the L1 Reflective Blogs

Also following the pilot study, the error analysis collected data using an error code schedule (see

Appendix F) designed to promote learning. Its development was in line with a schedule used in

research conducted by Ferris (2006) in that it was compiled, shared and refined over several

years by practising teachers based on the most frequent errors typically committed by students at

the particular level of learning (NCEA Level 1) and in the particular language (FFL) of the

research context. In addition, it was also approved in multiple annual NCEA assessment

moderation processes. Ferris' schedule contained fifteen codes but she recommended using

fewer. The schedule used in the current research comprised eleven codes, enough to incorporate

all the error types made by the students in the current research. These were placed in the margin

of the line where the error occurred so participants had to first find the error before making any

changes to their work. If participants were unable to find and fix an error, they could refer back to the coding schedule, which provided not only a definition of each code but also asked directional questions and indicated practice exercises specific to the error type. Graham and Macaro (2007) similarly offered the students in their research a choice of remedial practice exercises prior to rewriting. The categories and these scaffolding exercises comply with Assessment *for* Learning principles but are general enough that any improvements remain the student's work and therefore render the writing appropriate for NCEA assessment purposes and for the purposes of this research.

In addition to the CAF and error analyses, the L2 drafts were measured using the expert judgment tool, East's (2008) Scoring Rubric. This analytic measure of learning was chosen because it was research-informed and research-tested (East, 2009). In addition, it was designed to measure the L2 writing of students of German sitting various pre-university examinations. It could therefore be easily adapted to measure the writing of students of French sitting Level 1 NCEA writing standards. The rubric measured L2 written proficiency and provided information regarding overall quality of writing, not unlike the NCEA rubric, but, because of its analytic nature, it was able to measure the discrete changes that took place between the first and second drafts in a way that the NCEA rubric could not. The East rubric analyses language according to five sets of descriptors, each with seven levels, to arrive at a score out of 35 (see Appendix G). The five categories relate to content organisation, lexical competence, grammatical competence, mechanical competence, and intercultural understanding.

The second expert judgment tool used to measure fine-grained changes in the L2 writing was a rubric developed from Aljaafreh and Lantolf's (1994) Regulatory Scale (see Appendix H). It measured the microgenetic development of a specific language feature (in this case, NZC Level 6 verb knowledge) in three developmental stages and five levels. The developmental stages range from the intermental stage of other-regulation where learners rely on others to help them perform, to the partially intramental stage of partial self-regulation, and to the fully intramental stage of full self-regulation. Level 1 is where the learner is not aware of the form or is not yet able to use the form and where the teacher must assume full responsibility for instructing the learner. Rather than providing corrective help, the teacher's task is to bring the target form into focus and begin co-constructing the ZPD with the learner. At Level 2, the learner still relies heavily on the teacher. Because there is some development, an opening is provided for the teacher and the learner to

begin negotiating the feedback process and for the learner to begin to progress toward self-regulation. Help tends to still be explicit. At Level 3, the learner understands the teacher's intervention and is able to react to any feedback offered. The level of help required to correct the error starts to move the learner towards the strategic and implicit end of the regulatory scale. At Level 4, the learner begins to assume full responsibility for error correction and may even reject feedback from the teacher if unsolicited. Development has not yet become fully intramental or automatised because the learner still produces the target form incorrectly and may still need the teacher to confirm the correctness of any changes. At Level 5, the learner becomes more consistent in using the target structure correctly and across contexts. In most cases, use of the correct form is self-generated and automatised, the learner being fully self-regulated (Aljaafreh & Lantolf, 1994, p. 470).

These four measures, two objective and two judgment, were used to score the participants' first and second drafts according to L2 written proficiency, complexity, accuracy and fluency, verb error elimination and the development of self-regulated verb knowledge. Based on the sociocultural definitions in Sections 1.3.3 and 4.3.1., any positive changes between drafts were considered to be L2 learning. The learning scores for each language feature were tallied in intervention and comparison groups and compared both intergroup and intragroup. Where the comparison revealed a statistically significant difference between groups, this was considered a possible result of the reflection intervention. To provide a finer grained picture of the impact of the intervention, the average scores of intervention and comparison NCEA groups were also compared so the groups that benefitted the most and the least were revealed.

The L1 blogs that formed the reflection intervention provided further insights and explanations for the growth or lack of growth in L2 learning after the reflection intervention. To begin with, the blogs were collated and organised into NCEA grade groups using Word and OneNote. (The grades used were those awarded in the units when they reflected.) Then the blogs were analysed linguistically using Wordsift 2. The ability of these digital tools to count, reorder and collate data provided some clarity and produced patterns and themes for the later coding process and enabled more objectivity and less researcher involvement in the initial stages of analysis while retaining an element of creativity in the latter stages (Ellis & Barkhuizen, 2005).

An adapted version of Biggs and Collis' (1982) SOLO Taxonomy, a rubric requiring expert judgment, was also used to measure the depth of thinking in the blogs (see Appendix K). Level 1 indicated that the thinking was judged to be incompetent because the writer had missed the point or was passive or unengaged. Level 2 indicated that the thinking was unistructural because each section contained single relevant ideas or the writer had a superficial awareness evidenced in vague or fuzzy thinking. Level 3 indicated that the thinking was multistructural; the blog listed several relevant but independent thoughts that, while specific, lacked detail. Level 4 indicated the thinking was relational; the writer integrated several ideas by analysing, comparing and contrasting, criticising, explaining causes or justifying. Their thinking was specific and detailed, aligning with Oxford's (2011) strategy of conceptualising with details. Level 5 indicated the thinking was extended abstract; the writer was able to generalise ideas and apply them to new domains by hypothesising or theorising with specifics as well as details, which aligns with Oxford's strategies of conceptualising broadly and going beyond the immediate data.

Then, guided by Zimmerman's (1990) Self-regulation Strategies, Graham and Macaro's (2007) Formulation Strategies, and Oxford's (2011) Strategic Self-regulated Model of L2 Learning, a list of 43 different strategies was compiled (see Appendix J). The list was used to guide the coding of the strategies and goals reported in the blogs during each of the three writing processes of the research period: planning, composing, and reviewing.

After triangulating the collection and scoring of both the L2 writing and the L1 reflection data by using the quadrant of tools, the findings were summarised. Displaying the summaries visually in figures and tables helped clarify the results (Ellis & Barkhuizen, 2005) of the four L2 writing investigations and the findings of the six sets of insights from the L1 reflective blogs.

In order to establish the reliability of the scores that resulted from using the measures in the quadrant, two colleagues were employed, one an L1 French language assistant and the other a trained and experienced secondary school L2 French teacher. According to Graham, et al. (2012), in education contexts "[t]here are three common indices of inter-rater agreement: the percentage of absolute agreement, various versions of Cohen's kappa, and the intra-class correlation coefficient" (p. 7). Graham, et al. explain that the percentage of absolute agreement measures how often raters agree and is the best tool to use when many raters agree on a single point or when two raters agree frequently. However, this method does not account for chance

agreement or for consistent disagreement as Cohen's kappa is able to do. Cohen's kappa also calculates how often raters agree and adjusts for chance agreement. By contrast, intra-class correlation is most useful when measuring with rubrics or continuous scales as it calculates the proportion of variation that results from the person being rated as well as disagreement between raters (Graham, et al., 2012).

Graham, et al. (2012) suggested benchmarks to determine whether inter-rater agreement was sufficient or not. When using the percentage of absolute agreement, they suggested that there should be a minimum of 75% agreement with 90% being high agreement, and if the rates are more than one point apart, other tools should be used. For Cohen's kappa they suggested a minimum of .61 and a high of .81. However, they cautioned that this measurement tool depends on distribution, so if many students' grades are too similar, this tool will not show high agreement. When measuring for intra-class correlation, they suggested a minimum of .8 and a high of .9. Again, because this tool depends on variation in grades, it will not show high agreement if there is little variation in students' grades. Graham, et al.'s recommendations were followed in this study because they represent common practice in education contexts.

For the error analysis and the error-free clauses for the CAF analysis, the L1 French Language Assistant and I marked and coded the errors in all four scripts for all 71 participants. For the NCEA groupings, the number of clauses and dependent clauses for the CAF analysis, and the scores for the East (2008) analysis, seven participants (10%) were randomly selected. Using Microsoft Excel (Version 14.0.7173.5000, Microsoft Office Home and Student, 2010), the 71 participants were placed into alphabetical order according to family name and every tenth name selected. Then, after some training and conferencing, the L2 French teacher and I scored the four drafts per participant, meeting separately on three more occasions to further clarify the process.

To establish the reliability of the NCEA grades, the number of clauses, dependent clauses and error-free clauses used in the CAF analysis, and the number of verbs for the error analysis, the percentage of absolute agreement, which asks how often raters agree and is arguably the most appropriate tool to use when two raters agree frequently, was used. For the NCEA grades, this resulted in a 93% absolute agreement. For the number of clauses, dependent clauses and error-free clauses per composition, and the number of verbs there was full (100%) agreement.

For the scores in the East (2008) analysis, the tallies of nominal agreement, spelling and verb error in the error analysis, and the levels in the microgenetic analysis, an intra-class correlation coefficient was calculated using the software package, SPSS Statistics (Version 23, IBM, 2015). Intra-class correlation is most useful when measuring with rubrics or continuous scales as it calculates the proportion of variation that results from the person being rated as well as disagreement between raters (Graham, et al., 2012). In the East Analysis, there was agreement on 131 of the 140 discrete marks allocated, five of the scripts having one mark difference and two of the scripts having a two mark difference. These differences were accepted and a correlation coefficient of .985 was calculated using SPSS. For the error and microgenetic analysis, the verb error coefficient was .993, the nominal agreement error coefficient was .999, and the spelling error coefficient was .989. All scores met the high correlation benchmarks described in Graham, et al. (2012).

4.3.6: Assumptions

There were four assumptions made when devising the research design. An initial assumption was that L2 writing indicates the participants' L2 knowledge and that L1 written reflection indicates the participants' thinking. While they might indicate achievement and the cognitive and metacognitive processes involved in L2 writing, it must be acknowledged that they provide incomplete snapshots only. A second assumption was that the L1 reflections would add insight into the gain or lack of gain in the L2 writing that resulted from the reflection intervention. A third assumption was that the tools chosen to measure learning in the L2 writing and insights in the L1 reflections did indeed measure what they set out to measure, and they produced reliable and valid data. A fourth assumption was that inter-rater reliability scores calculated with a 10% sampling of participants' scripts, a 100% L1 French speaker's error check, and SPSS computations were sufficient to establish the reliability of the L2 data and the statistical significance of any gains.

4.4: Summarising the Chapter

This chapter used insights gained from the literature review in Chapter Two, the Pilot Study in Chapter Three and the review of the three classroom studies in Section One of this chapter to arrive at a research design capable of providing a reliable and valid response to the research questions: 1. Does L1 written reflection increase L2 learning when adolescents write for portfolio

assessment? If so, what is the type and amount of increase and for whom? 2. What insights, if any, does the content of the first language reflections reveal concerning the results of the L2 data analysis? 3. Are there any links between the reflective self-regulatory practices reported in the reflections, achievement levels, and/or continued learning? 4. What further insights regarding language learning do the reflections provide?

The design chosen to answer these questions was a mixed method, counter-balanced, comparative study that emanated from both sociocultural and interactionist perspectives of SLA. It gathered data from the L2 writing and L1 written reflections produced by Year 11 FFL students (n=71) and measured the learning gains that took place between their first and second drafts when they reflected and when they did not. It measured these gains using a balance of objective tools and tools that required training and expert judgment.

The next chapter presents and analyses the results from the L2 writing. It begins by presenting participant profile data gathered from a survey conducted near the end of the academic year and after the research period had ended.

CHAPTER FIVE: COMPARING L2 GAINS

The previous chapter detailed the chosen research design: a mixed method, counter-balanced comparative study framed by sociocultural-interactionist theories of SLA. This design seemed the most appropriate to address the research questions:

1. Does L1 written reflection increase L2 learning when adolescents write for portfolio assessment? If so, what is the type and amount of increase and for whom?

2. What insights, if any, does the content of the L1 reflections reveal concerning the results of the L2 data analysis?

3. Are there any links between the reflective self-regulatory practices reported in the reflections, achievement levels, and/or continued learning?

4. What further insights regarding language learning do the reflections provide?

This chapter presents the first set of results that addressed the first of these four questions by comparing the L2 gain of the seventy-one Year 11 participants when they took part in the reflection intervention and when they did not. The data consisted of 284 FFL compositions, two drafts of two tasks per participant, written three months apart, and 71 reflective blog entries, one per participant, written during the two-stage reflection intervention. The counter-balanced research design meant that the Term 1 intervention group became the Term 2 comparison group and vice versa, which meant that "each participant serve(d) as his/her own control" (Martella, et al., 2013, p. 154) and ensured that no student was disadvantaged or advantaged in the high-stakes portfolio assessment context in which the research took place. This design also accounted for order effect.

In this chapter, the L2 writing samples are analysed while the L1 reflections are analysed in Chapter Six. The L2 writing analysis was carried out via four investigations, each using a different tool to measure a different L2 feature in the scripts. The chapter begins by revisiting the research context, the participants, and the tools used to measure the learning gains.

5.1: Revisiting Context, Participants and Measures

As previously described, the research took place in five Year 11 classrooms in four New Zealand coeducational secondary schools. These schools represented a spread on the socioeconomic scale: deciles 4, 6, 7 and 10, decile 1 being the lowest and decile 10 the highest.

In each classroom the choice of textbook and course documents indicated that various communicative methodologies were in daily use. The age of the majority of participants (58/71 or 82%) at the end of the research period was fifteen years. The group comprised 27 (38%) males and 44 (62%) females who were learning FFL at NZC Levels 5 and 6 (see Section 1.2.2) which align with Global Scale Levels A2 and A2+ of the CEFR for Languages (Council of Europe, 2001). Each participant took part in the reflection intervention for one of the two writing units. Two classes reflected in Term 1 (n = 27) and three classes in Term 2 (n = 44).

Because the L2 writing gathered for the research could also be submitted for NCEA internal assessment, an adapted NCEA rubric was used to score the first drafts of each task, and the grades produced, consistent with the NCEA marking scheme, were used to place participants into pre-test groups for a more fine-grained analysis. This NCEA rubric measures written language holistically and according to descriptors related to text-type, language complexity, accuracy and fluency, and the unhindered communication of information, ideas and opinions (see Appendix B). It allocates one of four grades: Not Achieved, Achieved, Achieved with Merit or Achieved with Excellence. Because there were no participants with clear Not Achieved grades (one or two in each unit were borderline Not Achieved-Achieved), the participants were placed into three pre-test groups, Achieved, Merit, and Excellence according to the first draft written in each unit. This meant that the pre-test groups for each writing unit were slightly different in number.

As explained in Chapter Four, in addition to the NCEA rubric, the L2 learning gains that took place between the participants' first and second drafts were measured using two objective tools: CAF analysis and error analysis, and two tools that required training and expert judgment: East's (2008) Scoring Rubric and a situated rubric created from Aljaafreh and Lantolf's (1994) Regulatory Scale. Each of the four tools measured a different aspect of L2 learning. East's analytic rubric was used to measure the core construct, L2 written proficiency, with scores for each facet used to generate an overall score. The CAF analysis measured language complexity, accuracy and fluency separately. The error analysis measured type and frequency of error and verb error elimination rates. Aljaafreh and Lantolf's (1994) Regulatory Scale measured self-regulated verb knowledge. All four generated scale variables.

After using these four measures to produce four sets of data for each of the 284 drafts, three digital statistical programmes and several statistical operations were used to compute the

differences between the language produced by intervention and comparison groups. Firstly, Excel (Version 14.0.7173.5000, Microsoft Office Home and Student, 2010) was used to compute the difference between the participants' first and second drafts with any positive difference considered to be L2 learning (see Sections 1.3.3 and 4.3.1). Secondly, SPSS Statistics (Version 23, IBM, 2015) was used to compute data reliability with intra-class correlation coefficients. This is arguably the most useful approach when measuring with rubrics or continuous scales as it calculates the proportion of variation that results from the person being rated as well as disagreement between raters (Graham, et al., 2012) (see also Section 4.3.5). (The percentage of absolute agreement, which calculates how often raters agree and is the appropriate tool to use when many raters agree on one point or when two raters agree frequently, was also used when there was high rater agreement.) The gains between drafts were then transferred from Excel to SPSS and independent samples t-tests computed. These resulted in five statistics: sample size (n), mean differences (M) between groups, standard deviations (SD), t-values (t), and probability scores (p). The alpha level was set at .05. The t-test statistic was chosen as the most straightforward means of comparing the mean gains between drafts of the intervention and comparison groups in the series of four investigations into different facets of performance. Because SPSS does not specify the benchmarks for effect size (d) between two means, an online Cohen's d calculator (www.campbellcollaboration.org) was also used, .2 being a small difference, .5 being a medium difference, and .8 being a large difference (Cohen, 1988). Where a Levene's test for homogeneity of variance indicated unequal variance, a t-test not assuming equal variance was reported and dot plots were used to discover why there was unequal variance. The four investigations undertaken in this chapter are presented in Sections 5.2 to 5.5 to follow. Where extracts of participants' L1 and L2 writing illustrate a point, pseudonyms have been used to ensure participant anonymity. Errors in extracts have not been corrected. It should also be noted that the participants were not aware of any of these measures of learning apart from the one that pertained to their assessment programme, the NCEA rubric.

5.2: Investigation One: L2 Written Proficiency

The first investigation analysed the data produced by the first of the four measures, East's (2008) Scoring Rubric (see Appendix G), an analytic tool designed to assess the core construct, L2 written proficiency. East's rubric is considered analytic because of the level of detail it gathers, scoring language according to five categories and seven levels within each category, then

bringing the section scores to a total out of 35 (5 x 7). The wording of East's rubric was adapted to fit the Year 11 French context of the current research and the tasks that elicited the data.

To illustrate how the East analysis was conducted, an extract of Anna's Task 1 Draft 2 is presented as an example of Excellence level writing produced by students in Term 1 of their Year 11 French programme. For Task 1, participants were asked to write an online profile in order to gain an e-pal. An explanation of how the writing was awarded 33/35 for L2 written proficiency follows. In Table 13, the sections of East's rubric that best describe Anna's writing are in bold type. The extract contains portions of the beginning, middle and end of Anna's composition. Errors are italicised.[11]

« Bonjour! Je m'appelle Anna. J'ai quatorze ans. Le 7 avril, c'est mon anniversaire. J'ai une petite sœur qui s'appelle Sophie. Elle est mignonne. J'habite à Auckland en banlieue... Auckland est la plus grande ville de *Nouvelle Zélande*... La semaine *dernier*, c'était l'anniversaire de ma sœur. Il y avait une grande fête chez moi. Elle a invité toutes ses copines chez moi. J'ai préparé un gâteau au chocolat et ma mère a gonflé des ballons….Mon correspondant idéal est *le* garçon. Il est branché, très sympa, sociable et *extroverti*. Il a seize ans. Je souhaite qu'il *a mêmes intérêts* que moi. Merci! À bientôt! »

(Hello! My name is Anna. I'm fourteen years old. My birthday is the 7th April. I have a little sister called Sophie. She's cute. I live in an Auckland suburb… Auckland is the biggest city in New Zealand… Last week was my sister's birthday. There was a big party at our house. She invited all her friends to our place. I made a chocolate cake and my mother blew up some balloons… My ideal e-pal is a boy. He's with-it, very kind, sociable and extroverted. He's sixteen. I want him to have the same interests as me. Thank you! Talk soon!)

In the category of content coverage and flow of ideas, Anna was awarded a six because she included six of the seven points in the instructions as well as opinions and her composition flowed well because of logical sequencing. Her knowledge of vocabulary and its functional and idiomatic use, her grammatical variety and sentence-level competence, and her spelling, punctuation and use of accents were all very good for a student at this level. In her entire composition she made

---

[11] While most errors she made in the extract above are italicised, one error was an omission, the definite article in *'les* mêmes intérêts'.

very few errors and the ones she made were of a minor nature or in structures not taught in Year 11, such as the subjunctive mood.

Because Anna included a wide variety of correct and complex (for Year 11) vocabulary and grammar structures, such as a feminine comparative adjective followed by the correct preposition, and verbs referring to the present, past and future, these three categories were each awarded a seven. Her intercultural understanding, awareness of the reader and of the text-type gained a six.

Table 13: An Extract from East's Analytic Scoring Rubric (see Appendix G)

	Content coverage and flow of ideas	Knowledge of vocabulary and its functional and idiomatic use	Grammatical variety and sentence-level competence	Spelling & punctuation, including accents	Intercultural understanding (register, awareness of reader and text-type)
7	• covers all 7 points in the instructions • includes opinions and reasons • very logical sequencing • writing 'flows'	**• very effective choice & usage of words / idioms / functions** **• there is solid mastery of word form** **• the meaning is not obscured**	**• very few errors and of a very minor nature**	**• very few errors of spelling, punctuation, accents** **• the meaning is not obscured**	• very aware of French reader e.g. Task 1: compares NZ and France and requests and describes e-pal. • very appropriate opening and closing
6	**• covers at least 6 points in the instructions** **• includes opinions and reasons** **• logical sequencing** **• writing generally ''flows'**	• effective choice & usage of words / idioms / functions • there is mastery of word form • the meaning is not obscured	• a few errors but generally of a minor nature	• few errors of spelling, punctuation, accents • the meaning is not obscured	**• generally aware of French reader e.g. Task 1: describes/gives opinion about NZ/town or 'translates' e.g. en seconde and requests, describes e-pal** **• good opening, closing**

Anna's choice of content and language indicated that she was aware of matters that might interest a French teenager reading her online profile and what might fulfil her purpose of finding an e-pal. However she did not make any explicit comparisons between New Zealand and France, which would have been appropriate in the context, hence the six.

5.2.1: L2 Written Proficiency per Intervention and Comparison Group

As illustrated by the above example, the investigation drawing on East's (2008) rubric resulted in a score out of 35 for each of the 284 scripts. To establish data reliability, a 10% random sampling of the scripts (see Section 4.3.5) was scored by an L2 French teacher and his scores compared with mine. An intra-class coefficient of .98 indicated high inter-rater reliability (Graham, et al., 2012). Intra-class correlation, which, as previously stated, is most useful when measuring with rubrics (Graham, et al., 2012), was used to calculate data reliability in this instance (rather than the percentage of absolute agreement) because two of the twenty-eight scripts scored by the

two raters produced a two mark difference, while five produced a one mark difference (see Section 4.3.5). Even so, there was agreement on 131 of the 140 discrete categories (five categories for each of the 28 scripts in the sample).

To test the hypothesis that L2 written proficiency increases after L1 written reflection, independent samples *t*-tests were used to compare the gains between drafts of the intervention (reflection) and comparison (no reflection) groups, firstly in Task 1 and secondly in Task 2. Thirdly, the data from the two tasks were combined and a paired samples *t*-test was conducted to analyse the difference in gain between the combined intervention groups and the combined comparison groups from both tasks. Table 14 presents the results, with mean values representing the mean gain in L2 written proficiency between drafts as measured by East's (2008) analytic rubric.

Table 14: Impact of Reflection on L2 Written Proficiency per Intervention and Comparison Group

Group	*n*	*M*	*SD*	*t*	*d*	*p*
Task 1 Intervention	27	3.78	3.00	2.49	.60	.02
Task 1 Comparison	44	2.25	1.56			
Task 2 Intervention	44	3.09	2.48	-.29	.07	.77
Task 2 Comparison	27	2.93	2.11			
Both Intervention Groups	71	3.35	2.69	1.98	.33	.05
Both Comparison Groups	71	2.51	1.80			

As indicated in Table 14, those who had opportunity to reflect while completing the Task 1 writing assessment gained significantly more between drafts than those who did not reflect (M = 3.78, t(69) = 2.45, p < .05, d = .60). This was not the case for Task 2.

When comparing the second draft gains in written proficiency between drafts of the combined intervention groups and the combined comparison groups, the paired samples *t*-test indicated that those who had the opportunity to reflect in either Task 1 or Task 2 improved in L2 written proficiency significantly more than those who did not reflect (M = .85, t(70) = 1.98, p = .05, d = .33).

When conducting the *t*-tests, Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for Task 2 but not for Task 1. Owing to this violated assumption in Task 1, a *t*-statistic not assuming homogeneity of variance is reported in Table 14.

A dot plot representation of the Task 1 intervention and comparison groups' mean gains between drafts in L2 written proficiency indicated why the assumption of homogeneity of variance was violated. There were several outliers in the intervention group (see Figure 3).

Figure 3: Dot Plot of the Task 1 Intervention and Comparison Groups' L2 Gains
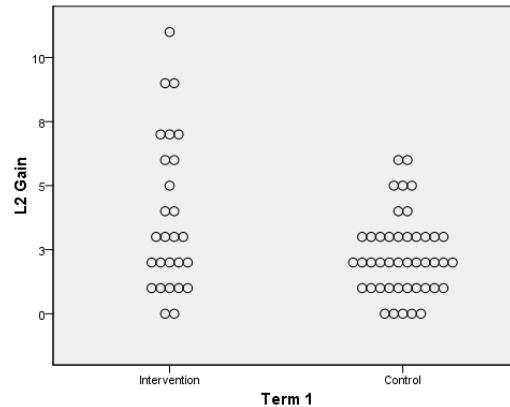
To explore the matter further, the East data were investigated in intervention and comparison NCEA subgroups.

5.2.2: L2 Written Proficiency per NCEA Subgroup

In this next stage of the East investigation, the intervention and comparison groups were divided into subgroups based on the pre-test NCEA grades awarded for each task, Achieved, Merit and Excellence. To establish the reliability of the grades awarded, the L2 French teacher scored a 10% sample of the scripts (28 in total). The percentage of absolute agreement was used in this instance due to both raters awarding all grades the same in Task 1 and all but one the same in Task 2. This produced 93% absolute agreement (see Section 4.3.5). To show the type of writing produced by the participants at each of the three NCEA levels, extracts from the Task 2 Draft 1 compositions are quoted below, with errors italicised. In Task 2, participants were asked to write a blog entry while travelling with an exchange partner.

 « *Apres* cela on a *prendu* le déjeuner et ensuite on a *fait tour de* sur la plage. Deux jours après ça, on a *faire de randonée dans le* "Broken Hills. » (Amelia, Achieved)

(After that we had lunch and then walked on the beach. Two days after that, we hiked at Broken Hills.)

 « Alors, hier, nous sommes allés à Auckland Zoo… Malheureusement, *il fait plu.* La météorologue *dit* que "Aujourd'hui, demain et *après demain fait* beau." Donc on est parti à treize heures. Je déteste *la météorologue.* » (Max, Merit)

(So, yesterday, we went to Auckland Zoo… Unfortunately, it rained. The forecaster said that today, tomorrow and the next day would be fine. So we left at 1pm. I hate forecasters.)

« Après avoir vécu en France toute sa vie, mon correspondant Pierre trouva la Nouvelle-Zélande un peu moins intéressant que la France. » (Edward, Excellence)

(After living in France all his life, my e-pal Pierre found New Zealand a little less interesting than France.)

To investigate the impact of the reflection intervention on the Achieved, Merit and Excellence subgroups when writing Task 1, independent samples *t*-tests were used to compare the mean gains in written proficiency between drafts of the intervention (reflection) and comparison (no reflection) groups within each grade group. Results are presented in Table 15, with mean values representing the mean gain in L2 written proficiency between drafts as measured by East's (2008) analytic rubric.

Table 15: Impact of Reflection on L2 Written Proficiency per NCEA Subgroup Task 1

Group		*n*	M	SD	*t*	*d*	*p*
Excellence	Intervention	9	1.44	1.24	-.06	.02	.96
	Comparison	17	1.47	1.07			
Merit	Intervention	14	4.93	2.90	2.91	.99	.01
	Comparison	22	2.50	1.47			
Achieved	Intervention	4	5.00	3.65	.51	.34	.64
	Comparison	5	4.00	1.58			

As indicated in Table 15, those in the Merit subgroup who had opportunities to reflect while completing the Task 1 writing assessment improved significantly more than those who did not reflect (M = 4.93, t(34) = 2.91, p < .05, d = .99). The effect size for this analysis (*d* = .99) was found to exceed Cohen's (1988) convention for a large effect (*d* = .80). The reflection intervention appeared to have little impact on the Task 1 Excellence and Achieved subgroups.

Figure 4: Dot Plot Showing Distribution of Task 1 Grade Groups' L2 Written Proficiency Gains
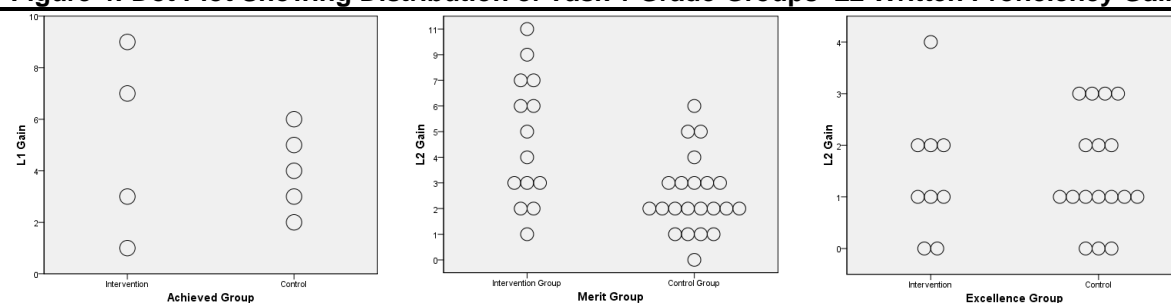
Levene's tests indicated that the assumption of equal variances was correct for the Task 1 Excellence subgroup but not for the Achieved and Merit subgroups. Owing to these violated

assumptions, a *t*-statistic not assuming homogeneity of variance is reported for the Achieved and Merit groups. Dot plots are also used to visually represent the variance (see Figure 4).

As indicated by the dot plots in Figure 4, the gain in written proficiency between drafts achieved by the small number in the Achieved intervention group (left) was polarised; and despite containing more participants and being not as polarised, the spread of the gain achieved by the Merit intervention participants (centre) was still wide compared with the comparison group. A similar spread with an outlier existed in the gains of Excellence intervention participants (right) but was more closely matched by the comparison group.

The analysis continued with the Task 2 data. To investigate the impact of the reflection intervention on the written proficiency of the Achieved, Merit and Excellence subgroups when writing Task 2, independent samples *t*-tests were used to compare the mean gain in written proficiency between drafts of the intervention (reflection) and comparison (no reflection) groups within each grade group. (The groups in this analysis were formed from the grades awarded the first draft of Task 2, and therefore contain a different set of participants from the Task 1 groups.)

Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for all subgroups in Task 2. Results are reported in Table 16. The mean values here represent the mean gain in L2 written proficiency between drafts as measured by East's (2008) analytic rubric.

Table 16: Impact of Reflection on L2 Written Proficiency per NCEA Subgroup Task 2

Group		*n*	M	SD	*t*	*d*	*p*
Excellence	Intervention	17	2.00	1.80	.18	.08	.86
	Comparison	7	1.86	1.68			
Merit	Intervention	16	3.19	2.64	-.11	.04	.92
	Comparison	18	3.28	2.70			
Achieved	Intervention	11	5.18	3.87	.41	.32	.69
	Comparison	2	4.00	1.41			

As indicated in Table 16, the reflection intervention appeared to have little impact on the written proficiency of the participants in Achieved, Merit and Excellence subgroups in Task 2. It may be speculated that for this task, the Term 1 intervention group, having reflected for Task 1, might have carried over the benefits of reflection when, as the comparison group, they wrote Task 2.

To investigate further, paired samples *t*-tests were conducted to compare the mean gains in L2 written proficiency between drafts of the combined Task 1 and Task 2 intervention (reflection) grade subgroups and the combined Task 1 and 2 comparison (no reflection) grade subgroups. Results are reported in Table 17, where the mean values represent the mean gains in L2 written proficiency between drafts as measured by East's (2008) analytic rubric.

Table 17: L2 Gains in Combined Term 1 and 2 Intervention and Comparison Groups

Combined Task 1 and 2 Grade Groups		*n*	M	SD	*t*	*d*	*p*
Excellence	Intervention Groups	26	1.81	1.63	.79	.36	.59
	Comparison Groups	24	1.58	1.25			
Merit	Intervention Groups	30	4.00	2.85	2.03	.49	.06
	Comparison Groups	40	2.85	1.89			
Achieved	Intervention Groups	15	5.13	3.68	.78	.15	.45
	Comparison Groups	7	4.00	1.41			

As indicated in Table 17, there was no statistically significant difference between the gains in written proficiency between drafts of the Task 1 and 2 intervention NCEA subgroups and the Task 1 and 2 comparison NCEA subgroups, the Excellence subgroups' *p*-statistic being .59 and the Achieved subgroups' being .45. However, the comparison of the mean gains in written proficiency between drafts of the combined Merit intervention subgroups did approach significance (*p* = .06).

Figure 5: Dot Plot of Term 1 and 2 Combined Merit Groups' Gains in Written Proficiency
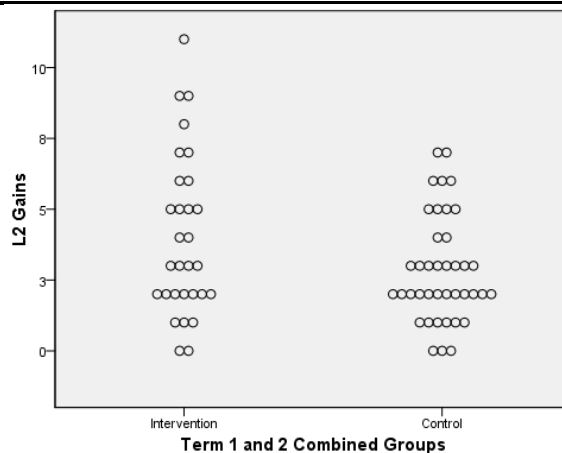
As indicated by the dot plot in Figure 5, the Term 1 and 2 combined Merit intervention groups' gains in written proficiency between drafts were widely spread and contained an outlier, as was seen previously when analysing the Task 1 Merit intervention group's gains.

5.2.3: Investigation One Summary

This first investigation tested the hypothesis that L1 written reflection increases L2 written proficiency when adolescents write for portfolio assessment. Independent samples t-tests indicated that there was a statistically significant difference between the Task 1 intervention and comparison groups' gains in written proficiency between drafts (p = .02) but there was no significant difference between the two groups in Task 2. When combining the intervention and comparison groups from both tasks, a paired samples t-test indicated that those who had the opportunity to reflect, whether in Task 1 or Task 2, gained in written proficiency between drafts significantly more than those who did not reflect (p = .05). When the study further investigated these results in intervention and comparison NCEA subgroups, it was found that the impact of the reflection intervention was statistically significant for the Task 1 Merit group but that it had little to no impact on other groups in Task 1 or on any group in Task 2. However, when combining the Task 1 and 2 Merit intervention (reflection) groups and the Task 1 and 2 Merit comparison (no reflection) groups, the gain in written proficiency between drafts of the combined intervention groups approached significance (p = .06).

5.3: Investigation Two: Complexity, Accuracy and Fluency

In the second investigation, the participants' drafts were measured using a set of CAF measures found in the pilot study to be suitable for measuring learning in writing produced by late beginners of a similar development level (see Chapter Three). The ratio of dependent clauses per clause (DC/C) was used to measure complexity, the ratio of error-free clauses per clause (EFC/C) to measure accuracy, and the number of error-free clauses per composition (EFC/Comp) to measure fluency. Unlike East's (2008) rubric, which was used in a way that aggregated its separate categories to one score, CAF analysis results in three stand-alone scores.

To introduce this investigation, clause, simple sentence, compound sentence and complex sentence are firstly defined. A clause consists of a subject and predicate which contains one finite verb. Simple sentences are those composed of one clause only. Compound sentences are composed of two or more clauses joined by coordinating conjunctions such as 'and, 'but' and 'therefore.' Complex sentences are those containing main clauses with dependent clauses joined by subordinating conjunctions such as 'because', 'if', and 'when.' Examples of a simple, a

compound, and a complex sentence from Ben's Task 1 Draft 2 illustrate this, with finite verbs underlined and coordinating and subordinating conjunctions italicised.

- Simple : « À la maison, j'<u>ai</u> deux chats. »

     (I have two cats at home.)

- Compound: « Ma ville <u>est </u>très touristique et pittoresque *mais* c'<u>est</u> ennuyeux. »

     (My town is very touristy and picturesque *but* it's boring.)

- Complex: « Je <u>suis</u> énervée avec ma sœur *parce qu'*elle <u>a mangé</u> tout le chocolat. »

     (I am annoyed with my sister *because* she ate all the chocolate.)

Before beginning the investigation, two further factors need to be noted, the overall simplicity of the participants' L2 writing and the small number of sentence-level changes that the participants made between drafts. Because complexity, in this investigation, is measured by subordination, the degree of simplicity was revealed by tallying the number of dependent clauses per NCEA subgroup.

As revealed in Table 18 below, for Achieved participants, the average Task 1 Draft 1 contained 29.75 clauses with only 1.75 (6%) dependent clauses. For Merit participants, their average Task 1 Draft 1 composition contained 33.36 clauses with 4.50 (13%) dependent clauses. For Excellence participants, while their compositions were longer, averaging 47.44 clauses, they wrote an average of only 7.0 (15%) dependent clauses.

Table 18: Increase in Number of Dependent Clauses per NCEA Group

	Group	Achieved		Merit		Excellence	
Task 1	Dependent clauses	1.75	6%	4.50	13%	7.0	15%
	Clauses	29.75		33.36		47.44	
Task 2	Dependent clauses	2.50	11%	3.43	15%	7.78	18%
	Clauses	22.83		27.43		43.44	
Increase	Task 1 to 2		5%		2%		3%

By mid-year, for Achieved students, the average Task 2 Draft 1 contained 22.83 clauses, 2.5 (11%) of them dependent, representing a 5% growth in complexity. For Merit students, of their average 27.43 clauses, 3.43 (13%) were dependent, representing a 2% growth. For Excellence students, of their 43.44 clauses, 7.78 (18%) were dependent, representing a 3% growth in complexity over the course of a term. By Term 2, the L2 writing of these Year 11 participants, therefore, still contained mainly simple and compound sentences rather than complex ones.

A comparison between the drafts produced by participants in intervention and comparison groups revealed that 39% of participants (28/71) in intervention groups and 46% (33/71) in comparison groups made no sentence-level changes at all between Draft 1 and Draft 2, focusing only on word-level changes. This resulted in minimal growth in complexity. In fact, sometimes changes in the second draft involved omitting error-laden complex sentences found in the first draft, thereby reducing rather than increasing complexity in the second draft.

Despite some growth in sentence-level complexity for all groups, by mid-year, the L2 writing of these Year 11 participants remained relatively unsophisticated. As a result, the third investigation supplemented this second CAF investigation with an analysis of verb usage (see Sections 5.4.1 and 5.4.4). This follows Ellis and Barkhuizen (2005) who stated that, for more elementary learners whose language contains relatively few dependent clauses (such as the participants in the present study), it behoves the researcher, when studying L2 learning, to supplement complexity measurement that uses subordination (as in the CAF analysis here) with an analysis of a specific linguistic feature, such as the verb. Knowing that verb analysis would follow, this second investigation went ahead and conducted the CAF analysis despite the small amount of growth between drafts and tasks.

5.3.1: Complexity, Accuracy and Fluency per Intervention and Comparison Group

To determine the reliability of data in this second investigation, CAF analysis, the L2 French teacher's tallies of the number of clauses, dependent clauses and error-free clauses in a 10% random sampling were compared with my tallies for the same group, and resulted in a 100% absolute agreement.

To test the hypothesis that complexity, accuracy and fluency increase after L1 written reflection, independent samples *t*-tests were used to compare the gains of the intervention (reflection) and comparison (no reflection) groups. Results are reported in Table 19.

Table 19: Impact of Reflection on CAF per Intervention and Comparison Group Task 1

	Group	*n*	M	SD	*t*	*d*	*p*
Complexity	Intervention	27	.00	.03	-.10	-.03	.92
	Comparison	44	.00	.03			
Accuracy	Intervention	27	.24	.16	1.17	.29	.25
	Comparison	44	.20	.11			
Fluency	Intervention	27	7.63	4.68	1.18	.29	.24
	Comparison	44	6.34	4.32			

As indicated by the results reported in Table 19, there was no statistically significant impact on complexity, accuracy or fluency in Task 1. The mean values here represent the mean gains in complexity, accuracy and fluency between drafts as measured by the ratio of dependent clauses per clause for complexity, the ratio of error-free clauses per clause for accuracy and the number of error-free clauses per composition for fluency.

Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for the Task 1 complexity and fluency analyses but not for the Task 1 accuracy analysis. Owing to violated assumptions in the Task 1 accuracy, a *t*-statistic not assuming homogeneity of variance is reported.

The investigation continued with the Task 2 data. To test the hypothesis that complexity, accuracy and fluency increase after L1 reflection, independent *t*-tests were used to compare the gains of the intervention (reflection) and comparison (no reflection) groups. The results are reported in Table 20. The mean values here represent the mean gains in complexity, accuracy and fluency between drafts as measured by the ratio of dependent clauses per clause for complexity, the ratio of error-free clauses per clause for accuracy and the number of error-free clauses per composition for fluency.

Table 20: Impact of Reflection on CAF per Intervention and Comparison Group Task 2

	Group	*n*	M	SD	*t*	*d*	*p*
Complexity	Intervention	44	.00	.02	-1.42	-.35	.16
	Comparison	27	.01	.03			
Accuracy	Intervention	44	.21	.13	-.76	-.19	.45
	Comparison	27	.24	.15			
Fluency	Intervention	44	6.34	4.32	-1.18	-.29	.24
	Comparison	27	7.63	4.68			

As indicated in Table 20, there was no statistically significant difference in complexity, accuracy or fluency between the intervention (reflection) and comparison (no reflection) groups in Task 2.

Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for the Task 2 accuracy and fluency analyses but not for the complexity analysis. Owing to violated assumptions, a *t*-statistic not assuming homogeneity of variance is reported for complexity.

To further the analysis, the data from the two tasks were combined and a paired samples *t*-test was conducted to analyse the difference in gain between the combined Term 1 and 2 intervention

(reflection) groups and the combined Term 1 and 2 comparison (no reflection) groups. Results are reported in Table 21.

Table 21: Impact of Reflection on CAF in Both Intervention & Both Comparison Groups

	Group	*n*	M	SD	*t*	*d*	*p*
Complexity	Intervention	71	.00	.03	-.79	-.19	.43
	Comparison	71	.01	.03			
Accuracy	Intervention	71	.22	.14	.49	.12	.63
	Comparison	71	.21	.13			
Fluency	Intervention	71	7.63	5.21	.04	.01	.97
	Comparison	71	7.61	4.70			

The mean values reported in Table 21 represent the mean gains in complexity, accuracy and fluency between drafts as measured by the ratio of dependent clauses per clause for complexity, the ratio of error-free clauses per clause for accuracy and the number of error-free clauses per composition for fluency.

The *t*-test indicated that those who had the opportunity to reflect in either Task 1 or Task 2 did not improve in complexity, accuracy or fluency significantly more than those who did not reflect. To further explore these results, the investigation continues in the next section with an analysis in intervention and comparison NCEA subgroups.

5.3.2: Complexity, Accuracy and Fluency per NCEA Subgroup Task 1

In this section, complexity in the Task 1 scripts is firstly analysed, then accuracy, followed by fluency. To test the hypothesis that **complexity** increases after L1 written reflection, independent samples *t*-tests were used to compare the mean gains in complexity between drafts of the Achieved, Merit and Excellence intervention (reflection) and comparison (no reflection) subgroups in Task 1. The Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for the Achieved, Merit and Excellence subgroups' complexity calculations. Results are reported in Table 22. The mean values here represent the mean gains in complexity between drafts as measured by the ratio of dependent clauses per clause.

Table 22: Impact of Reflection on L2 Complexity per NCEA Subgroup Task 1

	Group	*n*	M	SD	*t*	*d*	*p*
Excellence	Intervention	9	-.00	.01	-1.21	-.50	.24
	Comparison	17	.00	.01			
Merit	Intervention	14	-.01	.03	-.77	-.26	.45
	Comparison	22	-.00	.04			
Achieved	Intervention	4	.05	.05	1.45	.97	.19
	Comparison	5	.01	.03			

As indicated in Table 22, the Task 1 complexity analysis indicated that, for all three NCEA subgroups, the reflection intervention had no statistically significant impact.

To test the hypothesis that **accuracy** increases after L1 written reflection in Task 1, independent samples *t*-tests were used to compare the mean gains in accuracy between drafts of the Achieved, Merit and Excellence intervention (reflection) and comparison (no reflection) groups. The Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for the Achieved, Merit and Excellence subgroups' accuracy calculations. Results are reported in Table 23. The mean values in Table 23 represent the mean gains in accuracy between drafts as measured by the ratio of error-free clauses per clause.

Table 23: Impact of Reflection on L2 Accuracy in NCEA Subgroups Task 1

Group		*n*	M	SD	*t*	*d*	*p*
Excellence	Intervention	9	.13	.06	-1.26	-.52	.22
	Comparison	17	.19	.12			
Merit	Intervention	14	.31	.17	2.11	.72	.04
	Comparison	22	.21	.11			
Achieved	Intervention	4	.23	.14	.57	.38	.59
	Comparison	5	.17	.13			

As indicated in Table 23, those in the Merit subgroup who had the opportunity to reflect while completing the Task 1 writing assessment performed significantly better than those who did not reflect (M = .31, t(34) = 2.11, p < .05, d = .72). This was not the case for the Achieved and the Excellence subgroups. The effect size for the Merit subgroups (*p* = .72) approached Cohen's (1988) convention for large effect (*d* = .80).

To test the hypothesis that **fluency** increases after L1 written reflection in Task 1, independent samples *t*-tests were used to compare the mean gains in fluency between drafts of the Achieved, Merit and Excellence intervention (reflection) and comparison (no reflection) groups. Results are reported in Table 24. The mean values here represent the mean gains in fluency between drafts as measured by the number of error-free clauses per composition.

Table 24: Impact of Reflection on L2 Fluency in NCEA Subgroups Task 1

Group		*n*	M	SD	*t*	*d*	*p*
Excellence	Intervention	9	29.22	12.91	.85	.35	.42
	Comparison	17	25.35	6.46			
Merit	Intervention	14	11.57	4.55	-1.31	-.45	.20
	Comparison	22	13.95	5.76			
Achieved	Intervention	4	5.5	1.29	.22	.15	.83
	Comparison	5	5.2	2.39			

As indicated in Table 24, the Task 1 fluency analysis found that, for all three NCEA subgroups, the reflection intervention had no statistically significant impact.

Because Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for the Achieved and Merit subgroups but not the Excellence subgroup, a *t*-statistic not assuming homogeneity of variance is reported for the Excellence subgroup.

5.3.3: Complexity, Accuracy and Fluency per NCEA Subgroup Task 2

This section continues the CAF investigation by examining the impact of the reflection intervention on complexity, accuracy and fluency in Task 2. To test the hypothesis that **complexity** increases after L1 written reflection in Task 2, independent samples *t*-tests were used to compare the mean gains in complexity between drafts of the intervention (reflection) and comparison (no reflection) groups. The results are reported in Table 25. The mean values here represent the mean gains in complexity between drafts as measured by the ratio of dependent clauses per clause.

Table 25: Impact of Reflection on L2 Complexity in NCEA Subgroups Task 2

Group		*n*	M	SD	*t*	*d*	*p*
Excellence	Intervention	17	.01	.01	.47	.21	.65
	Comparison	7	.00	.02			
Merit	Intervention	16	.00	.03	-1.38	-1.16	.18
	Comparison	18	.02	.03			
Achieved	Intervention	11	-.01	.02	-.38	-.22	.71
	Comparison	2	.00	.00			

As indicated in Table 25, the Task 2 complexity analysis found that the reflection intervention had no statistically significant impact. Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for the Achieved, Merit and Excellence subgroups' complexity calculations in Task 2.

Table 26: Impact of Reflection on L2 Accuracy in NCEA Subgroups Task 2

Group		*n*	Mean	SD	*t*	*d*	*p*
Excellence	Intervention	17	.23	.13	-.89	-.40	.38
	Comparison	7	.29	.21			
Merit	Intervention	16	.21	.14	-.38	-.13	.71
	Comparison	18	.23	.13			
Achieved	Intervention	11	.18	.13	.36	.27	.73
	Comparison	2	.15	.00			

To test the hypothesis that **accuracy** increases after L1 written reflection in Task 2, independent samples *t*-tests were used to compare the mean gains in accuracy between drafts of the

intervention (reflection) and comparison (no reflection) groups in Task 2. The results are reported in Table 26. The mean values here represent the mean gains in accuracy between drafts as measured by the ratio of error-free clauses per clause.

As reported in Table 26, the Task 2 accuracy analysis found that the reflection intervention had no statistically significant impact. Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for the Achieved, Merit and Excellence subgroups' accuracy calculations in Task 2.

To test the hypothesis that **fluency** increases after L1 written reflection in Task 2, independent samples *t*-tests were used to compare the mean gains in fluency between drafts of the intervention (reflection) and comparison (no reflection) groups. Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for the Achieved, Merit and Excellence subgroups' fluency calculations in Task 2. Results are reported in Table 27. The mean values here represent the mean gains in fluency between drafts as measured by the number of error-free clauses per composition.

Table 27: Impact of Reflection on L2 Fluency in NCEA Subgroups Task 2

Group		*n*	M	SD	*t*	*d*	*p*
Excellence	Intervention	17	7.06	3.78	-2.87	-.09	.01
	Comparison	7	11.86	3.58			
Merit	Intervention	16	6.81	5.19	.23	.28	.82
	Comparison	18	6.44	4.25			
Achieved	Intervention	11	4.55	3.50	.41	.31	.69
	Comparison	2	3.50	0.71			

As indicated in Table 27, the Task 2 fluency analysis found that the reflection intervention had no statistically significant impact on the Merit and Achieved subgroups.

Figure 6: Dot Plots of Excellence and Achieved Subgroups' Gains in Term 2 Fluency
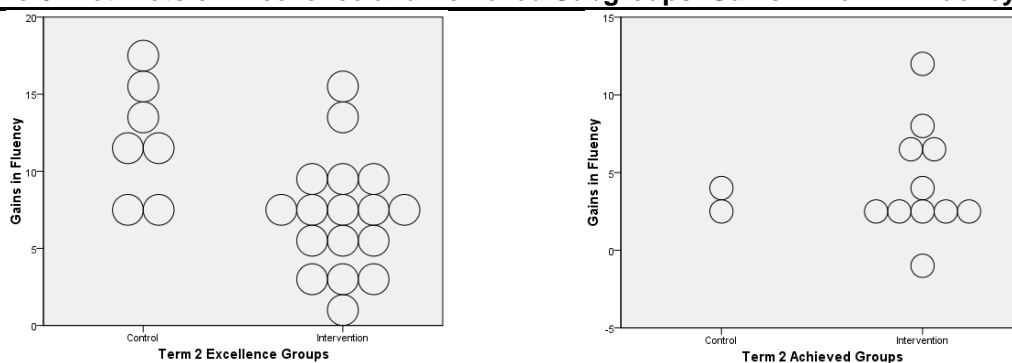
However, for the Excellence subgroup, there was a small but statistically significant negative effect (p = .01, d = -.09), perhaps because of the small group size and the unequal group sizes (see Figure 6). This was not the case for the Achieved subgroups, which were also small and unequal in size.

5.3.4: Investigation Two Summary

This second study used CAF analysis to investigate the impact of the reflection intervention on intervention and comparison groups and found that there was no statistically significant difference in complexity, accuracy or fluency either in Task 1 or Task 2, or when both intervention and comparison groups from Terms 1 and 2 were combined. When analysing intervention and comparison NCEA subgroups, it also appears that there was no significant positive difference in complexity, accuracy or fluency between the Task 1 and Task 2 Achieved and Excellence subgroups. However, there was a statistically significant difference in accuracy between the intervention and comparison Merit subgroups in Task 1 (but not in Task 2), and the effect size (d = .72) approached Cohen's (1988) convention for large effect (d = .80). There was also a small but significant negative effect on the Task 2 Excellence subgroups' fluency. This was possibly due to the small number in the Excellence comparison subgroup.

5.4: Investigation Three: Error Analysis

Having measured the impact of the reflection intervention on L2 written proficiency and on complexity, accuracy and fluency, the third investigation measures its impact on the elimination of error. The data for the investigation were generated by using a schedule of eleven indirect error codes (see Appendix F).

Figure 7: Extract from Joanne's Task 2 Draft 1 with Indirect Error Coding

These codes were placed in the margin of the scripts on the line where an error occurred so participants had to locate the error before making any changes to their work. Figure 7 illustrates

an example of indirect coded feedback in the margin of a portion of Joanne's Task 2 Draft 1. It indicates errors of omission (^), nominal agreement (A), word order (WO), verb (V), usage or idiom (U), redundancy (+), and spelling (Sp).

When the first drafts with error codes were returned to participants, they were invited to tally their codes and focus on their two or three most frequent error types. This responded to Bitchener (2008) who found that targeted feedback on error was more effective in promoting L2 learning than feedback on every error. The error investigation began by establishing the three most frequent errors of the entire participant group.

5.4.1: Most Frequent Error Types

In order to find the three most frequent errors, the participants tallied their errors themselves and recorded their most frequent types. In addition, an L1 French rater tallied the errors in a random sample of twenty-eight scripts produced by ten percent of participants. An intra-class correlation, which calculates the proportion of variation that results from the person being rated as well as disagreement between raters (Graham, et al., 2012), found that the reliability coefficient for verb error was .993, for nominal agreement error was .999, and for spelling error was .989. All met Graham, et al.'s (2012) benchmark for high correlation. The error tally found that the three most frequent error types for the majority of participants were spelling (Sp), verbs (V) and nominal agreements (A). These are reported in Table 28.

Table 28: The Participants' Three Most Frequent Error Types

	Spelling	Nominal Agreements	Verbs
Task 1:1	86%	72%	69%
Task 2:1	69%	58%	90%

As indicated in Table 28, the tally of error types revealed that, in Task 1 Draft 1, 86% of participants included spelling errors amongst their three most frequent types, 72% included nominal agreements, and 69% included verbs. By Task 2 Draft 1, 90% of participants included verb errors in their three most frequent error types, with 69% including spelling, and 58% including nominal agreements. Ellis and Barkhuizen (2005) suggested supplementing complexity analysis of simple text with analysis of a specific feature, such as the verb. The increase in verb error in the Term 2 scripts (compared with decreases in spelling and nominal agreement errors) supported this suggestion.

5.4.2: Verb Error per Intervention and Comparison Group

The error investigation in this section analyses verb error elimination between intervention and comparison groups. The analysis begins by calculating the ratio of verb error per verb in first drafts then in second drafts and calculating the difference between the two.

To test the hypothesis that verb error elimination increases after L1 written reflection, independent samples *t*-tests were used to compare the mean gains in verb error elimination between drafts of the intervention (reflection) and comparison (no reflection) groups. In addition, paired samples *t*-tests were used to compare the mean gains in verb error elimination between drafts of the combined Task 1 and 2 intervention groups and the combined Task 1 and 2 comparison groups. Results are presented in Table 29. The mean values here represent the mean gains in verb error elimination between drafts as measured by the ratio of verb error per verb.

Table 29: Impact of Reflection on Verb Error per Intervention and Comparison Group

Group	*n*	M	SD	*t*	*d*	*p*
Task 1 Intervention	27	.07	.11	.21	.05	.84
Task 1 Comparison	44	.07	.07			
Task 2 Intervention	44	.98	.11	.52	.13	.60
Task 2 Comparison	27	.08	.12			
Both Intervention Groups	71	.09	.11	1.20	.29	.23
Both Comparison Groups	71	.07	.09			

As indicated in Table 29, there was no significant difference in verb error elimination between the Task 1 intervention and comparison groups (*p* = .84), between the Task 2 intervention and comparison groups (*p* = .60), or between the combined Term 1 and 2 intervention groups and combined Term 1 and 2 comparison groups (*p* = .23).

Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for Task 2 but not for Task 1. Owing to this violated assumption in Task 1, a *t*-statistic not assuming homogeneity of variance is reported. As in previous studies, the investigation continues with intervention and comparison NCEA subgroups.

5.4.3: Verb Error per NCEA Subgroup

In this section, the ratio of verb error per verb was calculated and compared in NCEA subgroups. To test the hypothesis that verb error elimination increases after L1 reflection in Task 1, independent samples *t*-tests were used to compare the mean gains in verb error elimination between drafts of the intervention (reflection) and comparison (no reflection) Achieved, Merit and

Excellence subgroups. Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for the Task 1 Achieved and Excellence subgroups but not for the Task 1 Merit subgroup so a *t*-statistic not assuming homogeneity of variance is reported for this group in Table 30. The mean values in Table 30 represent the mean gains in verb error elimination between drafts as measured by the ratio of verb error per verb.

Table 30: Impact of Reflection on Verb Error per NCEA Subgroup Task 1

Group		*n*	M	SD	*t*	*d*	*p*
Excellence	Intervention	9	.04	.05	-.359	-.15	.72
	Comparison	17	.05	.05			
Merit	Intervention	14	.20	.13	.721	.25	.48
	Comparison	22	.07	.66			
Achieved	Intervention	4	.06	.12	-.736	-.49	.49
	Comparison	5	.12	.14			

As indicated in Table 30, the verb error analysis found that the reflection intervention had no statistically significant impact on verb error elimination for any of the NCEA subgroups in Task 1.

To test the hypothesis that verb error elimination increases after L1 reflection in Task 2, independent samples *t*-tests were used to compare the mean gains in verb error elimination between drafts of the intervention (reflection) and comparison (no reflection) Achieved, Merit and Excellence subgroups. Results are presented in Table 31. The mean values here represent the mean gains in verb error elimination between drafts as measured by the ratio of verb error per verb.

Table 31: Impact of Reflection on Verb Error per NCEA Group Task 2

Group		*n*	M	SD	*t*	*d*	*p*
Excellence	Intervention	17	.08	.07	.10	.05	.92
	Comparison	7	.07	.11			
Merit	Intervention	16	.09	.10	.29	.10	.77
	Comparison	18	.08	.12			
Achieved	Intervention	11	.14	.15	-.22	-.17	.83
	Comparison	2	.17	.13			

As indicated in Table 31, the reflection intervention had no impact on verb error elimination for any of the NCEA subgroups in Task 2. Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for all *t*-tests in Task 2.

5.4.4. Investigation Three Summary

The third investigation began by establishing the three most frequent error categories in the participants' L2 writing and by selecting verb error for further investigation. When analysing the

impact of the reflection intervention on the elimination of verb error, there was no statistically significant difference between the Task 1 and Task 2 intervention and comparison groups, whichever way they were considered. Studying the data in intervention and comparison NCEA subgroups confirmed these results. The next section continues to use the verb error data but adds data on revision outcomes and data analysed by Aljaafreh and Lantolf's (1994) Regulatory Scale to measure the impact of the reflection intervention on the self-regulation (microgenetic development) of verb knowledge.

5.5: Investigation Four: Self-regulated Verb Knowledge

The results of the previous investigation suggest that L1 reflection had little impact on verb error elimination. This suggested that participants needed more explicit support as well as reflective practice in order to increase their verb error elimination rates. The fourth investigation tested this hypothesis by analysing the impact of the reflection intervention on the microgenetic development of verb knowledge. The investigation combined the number of correct verbs (rather than errors) produced by the participants with their responses to feedback on verb error, and awarded one of five microgenesis levels using a rubric created from Aljaafreh and Lantolf's (1994) Regulatory Scale (see Appendix H). At Level 1 of this scale, a learner has little awareness of a particular language feature and the teacher must assume full responsibility for drawing the learner's attention to the feature and familiarising the learner with its form and function. By Level 5, the learner is able to use the form consistently and in multiple contexts. When an error is made, the learner notices and corrects the error without intervention from others. The feature has therefore become automatised and the learner fully self-regulated.

This microgenesis investigation involved three steps. The first was to calculate the average ratio of correct verbs per verb in first drafts, which included corrections made under test conditions, and to award one of the five verb microgenesis levels based on these two factors. Verb errors corrected under test conditions without access to resources were included in this first calculation as they were deemed to be mistakes rather than errors and indicated the learner's verb knowledge to be closer to the self-regulation end of the rubric. (It must be acknowledged, however, that participants had only ten minutes of class time to correct errors under test conditions. They might therefore have been able to change more language if they had been allowed more time.)

The second step involved calculating the ratio of correct verbs per verb in second drafts, which included corrections made at home with resources, and then awarding a second self-regulation level. These at-home corrections were deemed to indicate the learner approaching the mid-point of the rubric whereas verb errors left unchanged, changed wrongly, or avoided indicated the learner nearer the other-regulation end of the rubric.

The third step involved computing the difference between the two microgenesis levels using Excel, and the difference between the mean gains in verb microgenesis between drafts of the intervention and comparison groups using SPSS.[12] To illustrate the types of verb error, the revisions, and the measurement of self-regulation in a participants' writing, extracts from an Achieved participant's Task 2 drafts are quoted below. Verb errors and corrections are underlined and other errors are italicised.

A comparison of the Draft 1 and 2 extracts shows that Suzanne corrected the first two past perfect verb errors as a result of the indirect coded feedback but used the wrong accent. She left the third verb, a present that should be a perfect, unchanged. She corrected the fourth irregular present tense verb but not the fifth and added an extra past perfect verb that was incorrect. However, she was able to correct the final complex past perfect reflexive, perhaps because she learned it in class as a formulaic expression, or chunk, from the teaching on common errors, one of the four forms of feedback offered.

> Draft 1: « *La meilleur fois* est quand nous <u>avons aller</u> au *cinèma*. Nous avons <u>regarder</u> le nouveau film 'Jurassic World'. J'y <u>vais</u> trois fois! Normalement nous <u>prendrons</u> la voiture ou <u>aller</u> à pied. Nous <u>avons beaucoup de amusant</u>. »

> (We watched the new film 'Jurassic World'. I've seen it three times. Normally we catch the bus or walk. We really enjoyed ourselves.)

> Draft 2: « *La fois mieux* est quand nous <u>sommes allès</u> au *cinèma.* Nous <u>avons regardè</u> le nouveau film 'Jurassic World'. J'y <u>vais</u> trois fois. Normalement nous prenons la voiture ou <u>aller</u> à pied, mais pour aller au *cinèma*, nous <u>sommes pris</u> le bus. Nous nous sommes bien amusés. »

---

[12] Collecting data using Excel then importing that data into SPSS for further analysis was recommended by Grande (2015) because Excel is an excellent platform for manipulating and formatting data, as well as catching and correcting errors.

Suzanne was awarded an Achieved pre-test grade for her Draft 1. She had produced 30 verbs, seventeen of which were incorrect, a ratio of .43 correct verbs per verb. Suzanne found and corrected one verb under test conditions and, although she indicated that she had found five others by underlining them with a straight or wavy line in the feedback lesson, she 'corrected' three of them wrongly and avoided correcting two by omitting them in the second draft. She was awarded a pre-test Level 2 microgenesis score because, although she indicated that she had found some errors based on feedback, she was not always able to correct them without help. In Draft 2, when she had access to resources at home, she was able to correct eight verb errors but left three unchanged. In addition, she added one new error. Her accuracy rate improved from .43 to .68, her self-regulated verb knowledge moving from Level 2 to Level 3. Level 3 on the microgenetic scale indicated that more than half of her verbs were correct and she was able to notice and correct over half of her verb errors when they were drawn to her attention. She needed other support to correct the rest.

After scoring the 284 scripts in this manner, the data were checked for inter-rater reliability in two ways. Intra-class consistency calculated for verb error indicated a coefficient of .993, meeting the benchmark for high reliability (Graham, et al., 2012). Self-regulation levels were awarded based on the ratio of correct verb per verb, 95-100% verb accuracy considered to be functioning at Level 5, 75-94% at Level 4, 50-74% at Level 3, 25-49% at Level 2 and 0-24% at Level 1. Due to high similarity, inter-rater reliability for the microgenetic scores was established using the percentage of absolute agreement, a tool suitable when two raters agree frequently. The scores were the same in 279 out of 284 instances which indicated 98% absolute agreement.

5.5.1: Verb Development per Intervention and Comparison Group

To test the hypothesis that microgenetic verb development increases after L1 written reflection, independent samples t-tests were used to compare the mean gains in verb microgenesis between drafts of the intervention (reflection) and comparison (no reflection) groups and a paired samples t-test was used to compare the mean gains in verb microgenesis between drafts of the combined Task 1 and 2 intervention groups and the combined Task 1 and 2 comparison groups. Results are presented in Table 32. The mean values here represent the mean gains in verb microgenesis between drafts as measured by the rubric developed from Aljaafreh and Lantolf's (1994) Regulatory Scale (see Appendix H).

Table 32: Impact of Reflection on Verb Microgenesis per Intervention and Comparison Group

Group	n	M	SD	t	d	p
Task 1 Intervention	27	.19	.56	.87	.02	.39
Task 1 Comparison	44	.07	.55			
Task 2 Intervention	44	.16	.53	.87	.21	.39
Task 2 Comparison	27	.04	.65			
Both Intervention Groups	71	.17	.53	1.27	.30	.21
Both Comparison Groups	71	.06	.58			

As indicated in Table 32, it appears that the reflection intervention had little impact on the microgenetic development of verb knowledge in either Task 1 ($p = .39$) or Task 2 ($p = .39$), or when the intervention groups and comparison groups from both tasks were combined ($p = .21$). Levene's test for homogeneity of variance indicated that the assumption of variance was correct for all three *t*-tests. The next section further analyses these results in intervention and comparison NCEA subgroups.

5.5.2: Verb Development per NCEA Subgroup

To further test the hypothesis that microgenetic verb development increases after L1 written reflection, independent samples *t*-tests were used to compare the mean gains in microgenetic verb development between drafts of the Task 1 intervention (reflection) and comparison (no reflection) NCEA subgroups. Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for all NCEA subgroups in Task 1. Results are presented in Table 33. The mean values here represent the mean gains in verb microgenesis between drafts as measured by the rubric developed from Aljaafreh and Lantolf's (1994) Regulatory Scale (see Appendix H).

Table 33: Impact of Reflection on Verb Microgenesis in NCEA Subgroups Task 1

Group		n	M	SD	t	d	p
Excellence	Intervention	9	.00	.00	-.41	-.17	.69
	Comparison	17	.06	.43			
Merit	Intervention	14	.29	.73	1.52	.52	.14
	Comparison	22	-.05	.58			
Achieved	Intervention	4	.25	.50	-.99	-.66	.36
	Comparison	5	.60	.55			

As indicated in Table 33, there was no statistically significant difference in microgenetic verb development between the Task 1 Achieved, Merit and Excellence intervention and comparison groups.

To test the hypothesis that microgenetic verb development increases after L1 written reflection in Task 2, independent samples *t*-tests were used to compare the mean gains in microgenetic verb

knowledge between drafts of the Task 2 intervention (reflection) and comparison (no reflection) NCEA subgroups. Results are presented in Table 34. The mean values here represent the mean gains in microgenetic verb development as measured by the rubric developed from Aljaafreh and Lantolf's (1994) regulatory scale (see Appendix H).

Table 34: Impact of Reflection on Verb Microgenesis per NCEA Group Task 2

Group		*n*	Mean	SD	*t*	*d*	*p*
Excellence	Intervention	17	.00	.49	.51	.23	.61
	Comparison	7	.12	.58			
Merit	Intervention	16	.13	.62	.83	.29	.41
	Comparison	18	-.06	.64			
Achieved	Intervention	11	.27	.47	-2.12	-1.63	.06
	Comparison	2	1.00	.00			

As indicated in Table 34, there was no statistically significant difference in microgenetic verb development between the Excellence, Merit and Achieved intervention and comparison groups in Task 2.

Levene's test for homogeneity of variance indicated that the assumption of equal variances was correct for the Merit and Excellence subgroups in Task 2 but not for the Achieved subgroups. Owing to this violated assumption, a *t*-statistic not assuming homogeneity of variance is reported for the Task 2 Achieved subgroup. Because the comparison of the Achieved intervention and comparison subgroups produced a *p*-statistic that approaches significance ($p = .06$), albeit negative, the dot plots in Figure 8 were used to analyse the reasons for their unequal variance.

Figure 8: Dot Plot of the Task 2 Achieved Groups' Gain in Self-Regulated Verb Knowledge
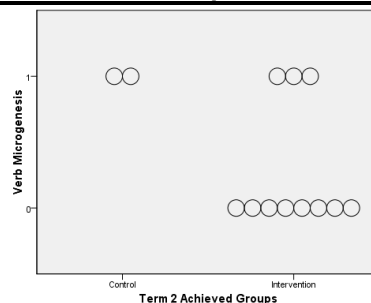
From the dot plot in Figure 8 it seems that the unequal variance resulted from the size disparity between the two groups and in particular from eight of the eleven participants in the intervention group not advancing a microgenetic level when both the participants in the comparison group did.

5.5.3: Investigation Four Summary

The investigation into the impact of the reflection intervention on the microgenetic development of L2 verb knowledge in the writing of the Year 11 participants found that there was no statistically significant difference either between intervention and comparison groups or between intervention and comparison NCEA subgroups in either Task 1 or Task 2, whichever way it was viewed.

5.6: Summarising the Chapter

This chapter set out to answer the first research question: 1. Does L1 written reflection increase L2 learning when adolescents write for portfolio assessment? If so, what is the type and amount of increase and for which students? Using four different measures of L2 learning, it conducted four investigations in an attempt to find answers (see Figure 9).

Figure 9: L2 Learning Gains after L1 Reflection per Investigation and Measure

Investigation Two:
Complexity, Accuracy, Fluency
Task 1 Merit Intervention Group's Accuracy

Investigation One:
Written Proficiency
Task 1 Intervention Group
Task 1 Merit Intervention Group
Combined Task 1 & 2 Intervention Groups

Objective Measures

Expert Judgment Measures

Investigation Three:
Verb Error Elimination
Little Impact

Investigation Four:
Microgenesis of Verb Knowledge
Little Impact

Investigation One used East's (2008) analytic rubric, which requires training and expert judgment, to measure the impact of the reflection intervention on the core construct, L2 written proficiency. Independent samples t-tests indicated that there was a statistically significant impact on the intervention group's written proficiency in Task 1 but not in Task 2. When the Task 1 and 2 intervention groups were combined and compared with the Task 1 and 2 comparison groups, paired samples t-tests indicated that there was again a statistically significant gain for the intervention groups. When the intervention and comparison groups were broken down into

intervention and comparison NCEA subgroups for further investigation, first the Term 1 groups and then the Term 2, independent samples t-tests revealed that the impact of the reflection intervention was statistically significant for the Task 1 Merit subgroup only and that it had little to no impact on the other subgroups in either Task 1 or Task 2, although the Task 2 Merit group's p-statistic approached significance ($p = .06$). It is worth noting that the effect size for the Term 1 Merit intervention group ($d = .99$) exceeded Cohen's (1988) convention for a large effect ($d = .80$).

The second investigation measured the impact of the reflection intervention on complexity, accuracy and fluency in the participants' L2 writing. This investigation found that there was no statistically significant difference between the intervention and comparison groups in complexity, accuracy or fluency in Tasks 1 or 2. However, when investigating further in intervention and comparison NCEA subgroups, it appeared that there was a statistically significant difference in accuracy between the intervention and comparison Merit subgroups in Task 1 and the effect size ($p = .72$) approached Cohen's (1988) convention for large effect ($d = .80$). The results of Investigation Two aligned with the results of Investigation One. As accuracy is a component of the East rubric, this confirmed the reliability of both the objective accuracy measure and the East measure that relied on expert judgment.

The third investigation analysed the impact of the reflection intervention on verb error elimination. It found that there was no statistically significant difference between the Task 1 or Task 2 intervention and comparison groups. Analysing the data in intervention and comparison NCEA subgroups did not alter these results. Whereas the East and accuracy analyses resulted in statistically significant differences for the Task 1 Merit group, in particular, the verb error elimination analysis did not.

The fourth investigation, an analysis of the microgenetic development of verb knowledge was therefore conducted. It combined data on correct verb production with data on revision outcomes and measured gain with a rubric formed from Aljaafreh and Lantolf's (1994) Regulatory Scale. It appears that the reflection intervention had no statistically significant impact on the microgenetic development of verb knowledge, whichever way this was viewed, either in intervention and comparison groups, in intervention and comparison NCEA subgroups, or in combined Term 1 and Term 2 intervention groups and combined comparison groups. This confirmed the results of the verb error elimination investigation.

In answer to the first research question, therefore, the reflection intervention appears to have had a medium to large effect on the L2 written proficiency and on the accuracy of the Term 1 Merit participants, in particular. It appears to have had little effect on complexity, fluency, verb error elimination or verb self-regulation for any group. The research now turns to the participants' L1 reflective blogs for further insight into the potential impact of reflection on learning.

CHAPTER SIX: FINDING INSIGHTS IN THE L1 REFLECTIONS

The previous chapter analysed the four L2 drafts produced by each of the 71 Year 11 participants when they wrote in FFL for portfolio assessment. The goal was to explore whether those participants who were given the time and the occasion to reflect in English as they wrote in French had a greater gain in L2 learning between drafts than those who were not given this opportunity. Gains were measured using situated versions of East's (2008) Scoring Rubric, CAF analysis, error analysis, and Aljaafreh and Lantolf's (1994) Regulatory Scale. From the statistical analyses conducted, it was found that the Term 1 Merit intervention group, in particular, increased their L2 learning in a statistically significant manner, and that they did so in terms of increased L2 written proficiency and in terms of accuracy (but not in terms of complexity, fluency, verb error elimination or self-regulated verb knowledge).

When researching L2 learning, there are three types of evidence available to the researcher: non-linguistic performance data, such as ticking a box in response to a listening exercise; language production samples, either spoken or written; and verbal report, either spoken or written (Ellis & Barkhuizen, 2005). Because each of these types of evidence provides an incomplete picture of language learning when considered on its own, research reliability and validity are more likely to be reached if the types of evidence are used in combination. As an example, Ellis and Barkhuizen suggested supplementing language production with verbal report to provide richer explanations for research phenomena. In the current research, the L2 writing data analysed in the previous chapter were supplemented with written report in the form of the participants' L1 reflections in the hope of acquiring richer explanations for significant gains after the reflection intervention.

In this chapter, the L1 reflections that accompanied the L2 writing are analysed. They were written by the participants when they took part in the reflection intervention, either in Term 1 or in Term 2, and provide a complementary data source for considering the perceived value and impact of L1 reflection on L2 learning. By analysing the reflections, the remaining research questions are addressed: 2. What insights, if any, does the content of the L1 reflections reveal concerning the results of the L2 data analysis? 3. Are there any links between the reflective self-regulatory practices evidenced in the blogs, achievement levels, and/or continued learning? and 4. What further insights regarding language learning do the blogs provide?

To begin with, the use of the term 'blogs' needs an explanation. In this research, the reflections were referred to as blogs for two reasons. Firstly, they were similar to blogs in that they were digitally written reflections equivalent to a single blog entry; and the Google Docs platform where they were recorded provided a place and a pattern for students to continue recording their progress after the research ended, should they wish to do so. Secondly, when explaining the research process to the parents and students, the reflections were referred to as blogs in order to avoid the increasingly negative response that the word 'reflection' elicits from secondary school students who are annually required to reflect on goal setting with their form teachers in conjunction with student-teacher-parent conferences. Currently, the use of the word 'blog' does not seem to have the same negative connotations for secondary school adolescents in New Zealand.

Regardless of what the exercise is called, giving learners the opportunity to pause during a lesson and record their thoughts has the potential to activate the self-regulatory and cognitive processes that are conducive to increased learning (Schunk & Usher, 2013; Swain, 2006; Zimmerman, 1990). It must be acknowledged, however, that, like any self-report, reflective blogs may not provide accurate or complete insights into language learner thinking (Cohen & Macaro, 2007). In this chapter, when extracts from the reflective blogs are quoted, pseudonyms are used to guarantee anonymity. It should also be noted by the reader that blog entries are presented 'as is'. That is, errors in language have not been corrected.

Figure 10: Outline of Chapter Six (Guided by Zimmerman's Dimensions of Self-Regulation)

6.1: L1 Reflection Data Overview 1. Data Collection 2. Lexis 3. Depth of Thinking	6.2: Time Dimension 1. Making Time 2. Managing Time	6.3: Method Dimension 1. Planning Strategies 2. Composing Strategies
6.4: Motive Dimension 1. Planning, Composing and Reviewing Goals 2. Goals for Continued Learning	6.5: Physical & Social Dimensions 1. Physical Environment 2. Social Environment	6.6: Behaviour Dimension 1. Evaluating Planning, Composing, Reviewing 2. Evaluating Feedback 3. Evaluating Final Comments

Figure 10 provides an outline for the chapter. Section 6.1 begins with a descriptive overview of the L1 language found in the blogs. Sections 6.2 to 6.6 analyse the blogs through the lens of Zimmerman's dimensions of self-regulated learning: Time, Method, Motive, Physical and Social Environments, and Behaviour. The dimensions were used in the analysis of the reflective blogs

because self-regulation is a key component of sociocultural theory, where language development is defined as the learner requiring increasingly less assistance from others as language knowledge becomes increasingly more internalised. In addition, the connection between L2 written performance and self-regulated learning was confirmed through research by Teng and Zhang (2016) (see Section 2.1.1).

In each section below, the processes and strategies that the participants reported in their blogs are coded and analysed, guided by the list of strategies (see Appendix J) compiled from the self-regulation strategies researched by Zimmerman (1990), the L2 formulation strategies researched by Graham and Macaro (2007), and the self-regulation language learner strategies researched by Oxford (2011). They are also analysed according to the recursive and dynamic writing processes of planning (conceptual preparation), composing (linguistic formulation and physical production) and reviewing (monitoring, editing and rewriting) that are referred to by Flower and Hayes (1981), Macaro, et al. (2016), Paesani (2006), and Schoonen, et al. (2009).

6.1: Overview of the L1 Reflection Data

This first section provides an overview of the L1 reflection data. To begin with, the data collection and coding processes are reviewed. Then the lexical character of the L1 language is analysed using Wordsift 2 (www.wordsift.org). Lastly, the depth of thinking apparent in the L1 blogs is analysed using the SOLO Taxonomy (Biggs & Collis, 1982).

6.1.1: Collecting, Coding and Collating the L1 Data

The reflection intervention took place on two separate occasions. The first occasion was immediately after, and within the same hour as, participants wrote their first L2 draft under test conditions. On this first occasion, participants received a hard copy of the reflection template and were invited to reflect on their planning and composition processes while their memories were still fresh. The second occasion occurred a week later, immediately after, and within the same hour as, receiving feedback on the first draft and making initial edits based on the feedback offered. On this second occasion participants used networked computers to transfer their previously recorded notes to digital format and to record further thoughts in a Google Docs blog. By reflecting immediately after rather than during writing and editing, it was hoped that these adolescent, post-beginner language learners would not be distracted from the L2 writing task at hand, and their cognitive load would be minimised while their memories remained fresh.

The reflections were guided by a template with five sets of prompts (see Appendix E). On the day they wrote the first L2 draft, the participants were prompted to reflect on their planning and composing processes, to consider whether they were happy with them or not, and, based on these reflections, to record planning and composing goals for future writing events. In the feedback lesson, participants were prompted to tally and consider their three most frequent error types and to set goals for editing and rewriting; their opinions on the helpfulness of each type of feedback (indirect error codes, teaching on common errors, teacher's comments and an NCEA grade indicator) were also sought. To end the reflection, participants were prompted to add any further comments they wished to make regarding their writing processes and strategies.

Once collected, the L1 data were coded and collated according to the three NCEA grade groups, the grades having been awarded to the first drafts of the L2 writing units in which the participants reflected (27 from the Task 1 unit and 44 from the Task 2). The coding and analysis process was supported digitally by Word (Version 14.0.7173.5000, Microsoft Home and Student, 2010) and OneNote (Version 16.0.7127.1026, Microsoft Home and Student, 2013). In the next section, Wordsift 2 (www.wordsift.org) provides a statistical overview of the language used. Together these three digital applications brought an organising and objective element to the relatively subjective and inductive coding process.

6.1.2: Statistical Analysis of the L1 Data

Wordsift 2 is a free online text analysis tool suitable for analysing a corpus of documents. It was created by Hakuta and Wientjes between 2008 and 2009 (Hakuta, 2016) under grants from the Council of Great City Schools, San Francisco, and the Strategic Education Research Partnership, and in collaboration with Stanford University and teachers in the San Francisco Unified School District. The collaborators' original motivation was to find a digital resource for classroom teachers wanting to improve the academic English of students in their various curriculum areas (Roman, et al. 2009). It was used in this current research to provide a snapshot of the lexical character of the participants' L1 blogs. Only the seven most pertinent statistics are presented here. To provide a contrast, the language in the first two paragraphs of this chapter (noted as 'Introduction' in Table 35) has also been analysed.

The statistics presented in Table 35 revealed that the Achieved participants wrote, on average, fewer words in the allocated time (192) than the Merit (231) and Excellence participants (242).

While the number of words per sentence was similar for all three groups (18.7, 19.42, 19.48), their sentences were more than a third shorter than the sentences in the introduction (36.3 words).

Table 35: Lexical Description of L1 Blogs by Wordsift 2

	n	Word Count	Lexical Density	Unique Words	Syllables per Word	Sentence Length	Grade of Readers	Readability Score
Excellence	26	242	40.54%	29.27%	1.40	19.48	9.44	68.02
Merit	30	231	39.62%	24.56%	1.34	19.42	8.56	73.50
Achieved	15	192	40.52%	39.30%	1.36	18.70	8.56	73.58
Introduction		290	51%	59.66%	1.8	36.30	19.2	21.3

Lexical density, measured by the ratio of meaningful words (nouns, adjectives, verbs, and adverbs) per word was also similar across the three groups (40.52%, 39.62%, 40.54%) yet their lexical density was less dense than the introduction (51%). Lexical variety was indicated by the ratio of unique words (used only once) per word. Even though there was a wider range between the three groups (39.3%, 24.56%, 29.27%), the lexis of all three showed less variety than the lexis used in the introduction (59.66%). While the average number of syllables per word was similar for the three groups (1.36, 1.34, 1.4), their words contained fewer syllables than the words in the introduction (1.8).

To comfortably read and understand the three groups' reflections, readers need to have reached similar grades in the US school system (grade '8.56', '8.56', '9.44') yet to read and understand the language in the introduction to this chapter, readers need to have a much higher level of education, indicated as grade '19.2'. The readability score, which measures the ease with which a passage might be read (0 being the most complex and 100 being the simplest), was, once again, similar across the three groups (73.58, 73.5, 68.02) but quite different from the introduction (21.3).

The Wordsift 2 analysis indicated that, while the participants with the highest L2 grades tended to write more, and in longer sentences, and they tended to use language that was richer and more complex, in comparison with the writing in the opening section of this chapter, the L1 writing of all participants remained relatively unsophisticated.

6.1.3: Depth of Thinking in the L1 Data

This section further describes the character of the blogs by presenting a measurement of the depth of thinking displayed in them. The responses in the blogs were divided into five parts according to the five sets of prompts in the template. Using a situated version of Biggs and Collis' (1982) SOLO Taxonomy (see Appendix K), each of the five parts was scored from Level 1 to

Level 5 thinking and the scores averaged. If a blog scored an average Level 1 on the SOLO

Taxonomy, the thinking was judged to be incompetent because the writer had missed the point or

was passive or unengaged. At Level 3, the blog contained lists of relevant and specific but

independent thoughts that lacked in detail. By Level 5, the participant was able to generalise

ideas and go beyond the immediate to apply ideas to new domains (see Section 4.3.5). To

establish intra-rater reliability, the depth of thinking in a 10% random sampling of the L1 blogs was

rescored after a three month interval. Using intra-class correlation, which is arguably the most

appropriate when measuring with rubrics as it calculates the proportion of variation that results

from the person being rated as well as disagreement between raters, SPSS was used to compute

a reliability coefficient of .85 which meets the benchmark for intra-rater reliability (Graham, et al.,

2012).

Table 36: Depth of Thinking per NCEA Group (SOLO Taxonomy)

	n	Range	Minimum	Maximum	M	SD
All	71	2.60	2.00	4.60	3.54	.58
Excellence	26	1.80	2.80	4.60	3.77	.49
Merit	30	1.85	2.40	4.25	3.50	.50
Achieved	15	2.00	2.00	4.00	3.23	.72

As Table 36 indicates, the mean depth of thinking for the participant group was 3.54, the thinking

of the Excellence participants being slightly more sophisticated than that of their counterparts

(3.23, 3.50, 3.77). Just as the length and L1 language used in the blogs paralleled the

participants' L2 grades, so, it appears, did their depth of thinking. Participants with higher NCEA

grades were more likely to:

- explain specifically and in detail the strategies they used as they planned and composed;

- set specific and detailed goals for future writing that correlated with strengths and
  weaknesses in first drafts;

- provide specific and detailed reasons why each form of feedback helped them; and

- critique the task itself and/or offer a summarising or global statement.

When comparing the Term 1 and 2 Merit groups for depth of thinking, the Term 2 blogs indicated

deeper thinking than the Term 1 blogs. The Term 1 group had a mean of 3.3 where the Term 2

group had a mean of 3.68. The slightly lower level of thinking for the Term 1 Merit group seems to

parallel their slightly lower L2 achievement as measured in the first draft.

Having reviewed in this first section how the L1 data were collected and having described the L1 data in terms of lexis and depth of thinking, the next five sections analyse the content of the L1 data according to Zimmerman's (2013) dimensions of self-regulated learning (Schunk & Usher, 2013). The first is the dimension of Time.

6.2: The Time Dimension

Based on the argument that self-regulated learners know how to choose the right time to learn and know how much time to spend on learning (Schunk & Usher, 2013; Zimmerman, 1990), in this section the L1 blogs are analysed for references to *time* and *time management*. Although *making time* is not listed as a strategy by Zimmerman, he specified *time* as one of the dimensions of self-regulated learning with *time management* as a key process. Oxford (2011) also did not list *making time* as a strategy, but stated that almost everyone who is interested can learn an additional language if they have sufficient time and they employ appropriate strategies. Equally, Graham and Macaro (2007) did not list *making time* as a strategy, but found that learners need sufficient time to practise using language learning strategies before their use becomes automatic.

6.2.1: Making Time

Wordsift 2 indicated that *time* was the most mentioned word in the participants' planning reflections. The format of the four week writing units purposely made time for the participants to plan by giving them the task seven days in advance of the test. Graham and Macaro (2007) found that the planning stage of process writing was largely omitted by British adolescent learners of FFL, so these seven days gave the participants in the current research the time they needed to plan by consolidating their concurrent classroom learning, generating and organising ideas in response to the writing task, and retrieving the language they needed to express those ideas. With access to resources during the planning time, they also had the opportunity to evaluate and overcome any gaps between what they wanted to say and what they were able to say by consulting resources and completing practice activities, and thereby increasing their L2 learning in the process (Shintani, et al., 2014). It was also hoped that, with proper planning, the cognitive load of composing without resources under test conditions would be reduced (Ong, 2014).

Two hesitations to this planning format (see Section 4.3.2) included participants committing error-laden language to memory or even memorising whole scripts instead of learning the language they needed to write on the day of the test. Typically, those who engage in such surface

strategies as rote learning produce poorer test grades than those who engage in deeper processing strategies, such as evaluating and overcoming the gap (Oxford, 2011).

Despite all participants receiving task instructions exactly seven days prior to the writing test, the blog entries indicated a marked difference in the amount of planning time spent by the participants in each NCEA group; and while all but two participants mentioned spending time, not all were specific regarding the amount of time they spent. Of Excellence participants who did specify time (n = 13), their commitment ranged from "not much time" to "about a week," 27% of the complete group (n = 26) wishing they had spent more time. For example, Bryce, an Excellence participant, recorded this entry:

> In preparation for the draft, I completely wrote down what I was to write and memorised key words in the document. It took about a week to eliminate errors and polish it off.

Of the Achieved participants who specified an amount of time (n = 6), their commitment ranged from "no time" to "the night before," 40% of the complete group (n = 15) wishing they had spent more time. For example, Eliza, an Achieved participant, wrote:

> I didn't prepare very well for the test. I wrote a practice draft the night before, but I think that I should have done this earlier in the week so I had more time to improve and make corrections.

Of the Merit participants who specified (n = 11), their planning time ranged from "barely" to "three days before." Of the complete group (n = 30), 33% wished they had spent more time. When comparing the reflection data of the Term 1 and 2 Merit groups, 36% of the Term 1 group wished they had spent more time compared with 25% of the Term 2 group. For example, Emma, one of the Term 1 Merit participants, wrote in her blog:

> For my preparation I went through my books and looked at some of the words and phrases that I might need, I spent around an hour just looking through all of the work that I thought could help.

This analysis of *time* in the participants' blogs seemed to indicate that the amount of time spent approximated the NCEA grade received for the first draft, the Achieved group spending the least time and the Excellence group spending the most time.

6.2.2: Managing Time

As well as *making time*, self-regulated learners are able to *manage time*. They know the connection between the task, the time needed to complete it and the need to choose times to work when they are mentally alert (Schunk & Usher, 2013). Jonathan, an Excellence participant, indicated the importance of *managing time* better when he wrote concerning his frequent errors: "this does not surprise me since I was not paying much attention on the day of the assessment. I will make sure to get more sleep before my next draft."

Only three (15%) Excellence participants and two (7%) Merit participants indicated any awareness of *managing time*. Despite her script memorisation, Alexa, a Term 2 Merit participant, indicated her awareness by writing the following in her blog:

> I prepared my text three days before the assessment was due. I am pretty happy with the preparation. I had enough time to learn everything to look over it again. Spot mistakes. I used a dictionary I look up words and a grammar book to see if my grammar was correct. Every day I learned a third of the text off by heart. That worked quite well.

While all but two of the 71 participants mentioned *making time* and/or *managing time* when planning, no Achieved participants, only two Merit participants and four Excellence participants referred to time during the composition process (the test). Two of the Excellence participants indicated they were pleased with how they used their time. Rohan wrote: "My time management during the writing process was good as I finished without stress or pressure." Lance wrote: "I had plenty of time to read over my work and check for errors, which was good." Two Merit and two Excellence participants felt they could have *managed time* better. Brigitte wrote:

> I finished before the time, which showed that I was well prepared. To be honest, I could have spend more time on it, to have a third or fourth look over it, to be sure, that I didn't miss any mistakes. Next time I would use the same preparing strategie, but spend more time with the writing and correction in class.

Maddy wrote: "I defiantly need to evaluate my ideas a lot more, I also need to give myself lots of time to finish." Summer wrote: "I'm not surprised that I made lots of spelling mistakes because I didn't have time to proofread it." Monika wrote: "I used some incorrect verb endings and got accents wrong. I think some of this came down to time pressure."

When researching time spent on writing processes, Roca de Larios, et al. (2008) found that less proficient writers spent more time than more proficient writers converting ideas into language (formulating), and less time monitoring and evaluating their work. Because so few participants mentioned having time to monitor their writing during the composition process, it is difficult to draw any conclusions from the current research, but it could possibly indicate that the participants spent most of the thirty minutes formulating and had little time left for monitoring.

After researching the different amounts of time students spent on different processes, Sasaki (2000) was able to write a differentiated curriculum for three levels of writers: experts, novices before instruction and novices after instruction (see Section 2.1.1). These three groups seem to equate to Budoff's (1968) high scorers, gainers and non-gainers (Lantolf & Poehner, 2014) (see Section 1.3.3). They also seem to equate to the three groups in this research, the participants who gained Excellence in their first drafts, those who appeared to gain significantly from the intervention (the Merit group), and the less proficient (the Achieved group) who did not seem to benefit from the intervention at all.

This second section explores the L1 reflective blogs for references to *time*. It was found that *time spent* was a primary concern mentioned by participants in relation to their planning strategies. It seemed that, the higher the NCEA grade awarded for the L2 writing in the unit when the participants reflected, the more time they reported spending on planning. Even so, one third of participants in all NCEA groups regretted not spending enough time relative to the demands of the writing test. It was interesting to note that, the more time the participants spent on planning, the more they reported memorising content, language or scripts. Very few participants indicated their ability to *manage time* effectively, either while planning (only five) or while composing (only six).

6.3: The Method Dimension

While the Time dimension of self-regulated learning refers to the process of time management, the Method dimension addresses the way that self-regulated learners select, use and orchestrate strategies to achieve learning tasks. According to Rubin (2008), effective strategy use involves the learner selecting strategies that relate to the task at hand and that suit the learner's learning preferences, and it involves the learner using these effective strategies in combination. Rubin stated that: "it is the operationalisation of the strategy that is critical, not the strategy, in and of itself" (p. 12). In the current research, the reflection template invited participants to evaluate the strategies they used when planning and composing. Throughout their reflections, the participants mentioned or implied use of fourteen of the strategies listed in Appendix J which were compiled from Zimmerman (1990), Graham and Macaro (2007) and Oxford (2011). Using the categories devised by Oxford, the strategies used by the participants were coded according to meta-strategies (M), cognitive strategies (C), affective strategies (A) and sociocultural-interactive strategies (S):

1. record keeping (the reflective blog itself) (M)
2. organising (M)
3. brainstorming and researching content (C)
4. activating prior learning (language that matches the task requirements) (C)
5. reviewing (notes, books, tests) (C)
6. evaluating and overcoming the language gap (consulting resources, doing practice activities, translating) (C)
7. monitoring (prompting, backtranslating, attending to complexity, accuracy, variety, style) (C)
8. rehearsing (writing practice sentences, drafts) (C)
9. memorising/remembering (language, ideas, scripts) (C)
10. going beyond the immediate data (C)
11. conceptualising with details (C)
12. conceptualising broadly (C)
13. seeking social assistance (S)
14. activating supportive emotions, beliefs and attitudes (A).

The vast majority of the strategies used by the participants were cognitive and related directly to language learning, two were meta-strategies which related to organising resources and writing the reflection, one was affective and related to managing emotions, while one was sociocultural-interactive because it related to seeking help from or working with others. In the following

sections, the participants' use of these strategies is analysed according to their NCEA grade groups and according to the writing processes of planning and composing.

6.3.1: Planning Strategies

To begin the analysis, the planning strategies mentioned by two or more participants in each NCEA group were compared. The analysis found two similarities between the groups, as well as several differences. As indicated in Table 37, the first similarity was that the most frequently used planning strategy for all three groups was rehearsing, which most often took the form of practice sentences or drafts (60% of Achieved participants, 70% of Merit and 73% of Excellence). The second similarity was their use of the strategy, activating prior learning. It was used by a similar percentage in each group (20% of both the Merit and Achieved groups and 19% of the Excellence). While the reported use of the strategies of rehearsing and activating prior knowledge was similar, there were also several differences between the three groups.

Table 37: Planning Strategies Used by Percentage of NCEA Group

	n	76-100%	51-75%	26-50%	0-25%
E*	26		Rehearsing (73%)	Memorising (39%) Monitoring (35%) Overcoming gap (35%)	Reviewing (23%) Activating learning (19%) Brainstorming content (19%) Managing time (15%) Critiquing the task (8%)
M*	30		Rehearsing (70%)	Reviewing (33%) Overcoming gap (33%) Monitoring (30%)	Memorising (23%) Activating learning (20% Managing time (7%) Brainstorming content (7%)
A*	15		Rehearsing (60%)	Reviewing (47%)	Overcoming gap (20%) Activating learning (20%)

E* = Excellence Group, M* = Merit Group, A* = Achieved Group

The first was their different use of the memorising/remembering strategy. It appears that the more time participants spent when planning, the higher the NCEA grade they received on their first draft and the more likely they were to use the memorising/remembering strategy. Approximately a third (39%) of Excellence participants stated that, as part of their planning, they memorised difficult words and phrases, key ideas or their entire script (8%). Two Excellence participants explained why they chose not to memorise a script as a planning strategy. Amanda stated: "I prefer to have the skills to write on the spot, rather than pre-write something and memorise it." Jess was more graphic: "I didn't bother memorising it because I think that's a bit pointless. I would like to be able to write in French well on the spot as opposed to being able to vomit a paragraph that I memorised the night before." However, of the 23% of Merit participants who mentioned

memorising, all but one referred to memorising the entire script. One Achieved participant mentioned memorising "key points in French". While eight (11%) of the 71 participants mentioned memorising entire scripts, and ten (14%) mentioned memorising key language or key ideas, fifty-three participants (75%) did not mention memorising/remembering at all.

A second distinction between the groups was their reported use of reviewing, monitoring, and evaluating and overcoming the gap. While the strategy of reviewing is non-specific and refers to going over notes, books and tests, monitoring and evaluating and overcoming the language gap have particular language features in mind and indicate a more critical awareness of language. Monitoring uses tactics (such as, prompts, mnemonics, rules and backtranslating) to attend to language features known to be troublesome, while also checking for complexity, style and task fulfilment. Evaluating and overcoming the gap uses tactics such as, consulting resources, doing practice activities or translating. The analysis revealed that more Achieved participants mentioned reviewing (47% compared with 33% of Merit participants and 23% of Excellence); yet more Excellence participants mentioned monitoring (35% compared with 30% of Merit participants and 7% of Achieved); and more Excellence participants mentioned evaluating and overcoming the gap (47% compared with 33% of Merit participants and 20% of Achieved). It seems that the higher the grade received, the more likely it was that the participant used strategies that indicate critical sensitivity to language (Naiman, et al., 1978; Zimmerman, 1990) rather than the more general and non-specific strategy of reviewing.

The third aspect that distinguished the NCEA groups was the variety of strategies used by each group. The Achieved participants mentioned using half the number of strategies used by the other two groups. In addition to the already mentioned strategy of monitoring, only one Achieved participant mentioned brainstorming content (compared with 19% of Excellence and 7% of Merit) and no Achieved participant mentioned managing time as they planned to write (compared with 15% of Excellence and 7% of Merit). Only Excellence participants mentioned critiquing the task (8%).

The fourth difference was the manner in which each group combined or did not combine the strategies as they prepared to write. Oxford (2011) called this 'orchestrating strategy use'. As indicated in Figure 11, 46% of Excellence participants mentioned combining three or more planning strategies (31% used three and 15% used four or more) compared with 37% of the Merit

group (20% used three and 17% used four or more) and 14% of the Achieved (one used three and one used four or more). Orchestrating strategies also seemed to approximate grades. As Stephanie, an Excellence participant, planned for the writing test, she indicated her ability to orchestrate four strategies: brainstorming (ideas about the topic), rehearsing (by structuring paragraphs and finishing her draft), activating prior learning (by thinking about tenses and an appropriate opening and ending), and monitoring (by checking grammar with websites and a dictionary). She wrote: "I brainstormed ideas about the topic, structured paragraphs and thought about the tenses and opening and ending. When I finished the draft, I checked grammar mistakes using the website given by the teacher and dictionaries."

Figure 11: Strategy Combination Per NCEA Subgroup
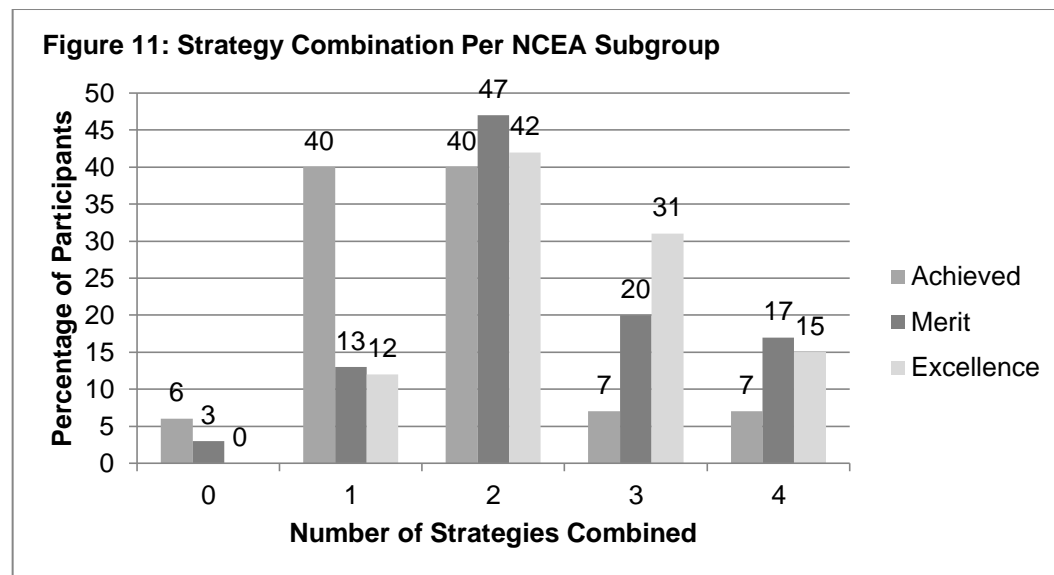
When comparing the strategy use of the Term 1 and Term 2 Merit intervention groups, a marked difference was revealed. Where 64% of the Term 1 Merit intervention participants mentioned the 'easier' strategy of reviewing, 19% of the Term 2 group mentioned reviewing. All other strategies were mentioned by more Term 2 Merit intervention participants than Term 1 – 81% compared with 57% mentioned rehearsing, 63% compared with 14% mentioned evaluating and overcoming the gap, 38% compared with 21% mentioned monitoring, 38% compared with 14% mentioned memorising, and 19% compared with 14% mentioned activating prior learning. There was also a marked difference in strategy orchestration between the Term 1 and 2 Merit intervention groups. Twenty-two percent of the Term 1 group combined three or more planning strategies compared with 50% of the Term 2 group. If strategy use is "strongly associated with superior academic functioning" (Zimmerman, 1990, p. 8), then it seems that the Term 1 Merit intervention group's L2 achievement as well as their strategy use might be inferior to the Term 2 Merit intervention

groups' L2 achievement and strategy use. It must be conceded that linguistic proficiency may well have developed between Terms 1 and 2 for all Merit participants by virtue of their engagement in teaching and learning beyond taking part in this study.

Revisiting the L2 data related to Task 1 Draft 1 confirmed this. Where the Term 1 group's writing contained an average of 33.36 clauses, 4.5 dependent clauses and 11.57 error-free clauses, the Term 2 group's contained an average of 38.95 clauses, 6.27 dependent clauses and 13.95 error-free clauses. Where the Term 1 group scored an average of 23.71 on East's (2008) L2 written proficiency scale, the Term 2 group scored an average of 24.09. Where the first drafts of the Term 1 group contained an average of 31.04 correct verbs out of 37.85, the first drafts of the Term 2 group contained an average of 32.05 out of 38.68. On every achievement count used in this study, the Term 1 Merit group scored lower in their first drafts than the Term 2 group. It seems that the Term 1 Merit group sat closer to the Achieved group on the NCEA achievement scale, while the Term 2 Merit group sat closer to the Excellence group. The Term 1 Merit group clearly had more scope for improvement both in L2 learning and in strategy use than their Term 2 counterparts.

6.3.2: Composing Strategies

After the seven day preparation period, the participants spent thirty minutes under test conditions and without resources composing their first drafts. Immediately after and in the same lesson as the test, the participants who took part in the reflection intervention were prompted to consider the difficulties they had experienced when composing and the strategies they used to overcome those difficulties. As indicated in Table 38, the most frequent difficulties reported by the 71 participants were remembering or thinking of key vocabulary and expressions (33.8%), knowing spelling and accents (29.6%), forming and linking sentences or sentence flow (18.3%), and knowing verb conjugations (16.9%).

Table 38: Difficulties when Composing per Percentage of NCEA Group

	n	Remembering vocabulary	Spelling and accents	Forming Sentences	Conjugating Verbs
All	71	34%	30%	18%	17%
Excellence	26	23%	42%	19%	19%
Merit	30	27%	27%	13%	13%
Achieved	15	67%	13%	27%	20%

The most reported difficulty for Excellence participants was spelling and accents (42%); for Achieved participants it was remembering vocabulary (67%), and for Merit participants it was remembering vocabulary (27%) and spelling and accents (27%). As is understandable, more composition difficulties were reported by the Achieved group. However, while all but two participants (97%) mentioned at least one difficulty when composing, just over half (56%) mentioned strategies to overcome those difficulties.

Table 39: Strategies for Overcoming Difficulties in Group Percentages

	n	Visual Monitoring	Prompting	Finding Alternatives	No Strategy
All	71	16%	14%	11%	44%
Excellence	26	19%	8%	4%	39%
Merit	30	10%	17%	10%	53%
Achieved	15	20%	20%	27%	33%

As indicated in Table 39, the strategies that 56% of participants used most frequently to overcome composition difficulties were visual monitoring (16%), where participants wrote troublesome language out multiple times to see what 'looked right', prompting or activating prior knowledge with rules or mnemonics (14%), and finding alternatives (11%). Other strategies mentioned by at least two participants included linguistic reasoning (6%), avoidance (4%), auditory monitoring (4%), restructuring (3%) and persevering (3%).

Of the four participants (6%) who applied reason to overcome their difficulties, two reported using "common sense" or "logic", and two reasoned based on knowledge transferred from other languages. For example, Bella, an Achieved participant, wrote: "I had a couple mind blanks for what certain words were in French but know them in Bulgarian and English. To overcome this, I tried to see if there was a root word within both translations and tried to fit it in a sentence, remembering what the French version would be using that root word."[13] Peta, a Merit participant wrote: "As I was writing in French, I had difficulties remembering certain nouns and verbs that I might need. For example, the word for cricket, so I guessed it and wrote le cricket which was in fact correct."

When comparing the participants' responses in NCEA groups there appeared to be differences. More participants from the Achieved group (67%) reported strategies to overcome difficulties than

---

[13] Given the 41% who indicated speaking a language other than English at home or at school, it is curious that more participants did not report similar reasoning.

the other two groups (61% of Excellence participants and 47% of Merit). It seems logical that

those who had the most difficulties would use the most strategies to overcome them, although this

runs counter to the idea that strategy use approximates achievement level (Cohen, 2011;

Graham, 2004; Schunk & Usher, 2013; Zimmerman, 1990).

In addition, the strategy mentioned most by Excellence participants was visual monitoring (19%),

by Merit participants it was prompting (17%), and by Achieved participants it was finding

alternatives (27%). However, the percentage of students who used visual monitoring in the

Excellence and Achieved groups was similar (19% and 20%).

Blog entries written by Lily, an Excellence participant, Priya, a Merit participant, and Alistair, an

Achieved participant, are representative of the three groups:

> I had some spelling troubles for example I wasn't sure if it was masculine or feminine so i
>
> mumbled it under my breath to see if it was M or F. Sometimes forgot how to write the
>
> more complex sentences. I visualised it in my head and muttered to myself. (Lily,
>
> Excellence)

> I had difficulty remembering verb agreements and verb genders, I overcame this by
>
> remember that in most cases verbs that end with 'E' are more likely to be feminine. Next
>
> time I will try use more new language. (Priya, Merit)

> If i couldn't think of a word in French i would use another word that meant the same thing or
>
> very similar to it. I think i could have used bigger describing words in there and a range of
>
> words instead of the same ones all the time, but i can always add that in. (Alistair,
>
> Achieved)

This section focused on the Method dimension of self-regulation and examined the way the

participants selected, used and orchestrated their strategy use while planning and composing in

FFL. It was found that, the higher the NCEA grade, the more effective the planning strategies

were, the greater the variety of strategies were used, and the more likely the strategies were to be

used in combination. The Excellence participants were also more likely to use planning strategies

that indicated critical sensitivity to language (monitoring, and evaluating and overcoming the gap).

Perhaps the most striking feature of this analysis was that, although 97% of participants

mentioned composition difficulties, only half (56%) indicated that they understood how to use strategies to overcome those difficulties.

6.4: The Motive Dimension

The next dimension of self-regulation to frame the analysis of the L1 reflections is the Motive dimension. This dimension is concerned with: why learners engage in learning and choose to continue learning; the goals they set to achieve learning that is important to them; and their subsequent self-efficacy, which is the belief that they have the capacity to achieve their goals (Schunk & Usher, 2013; Zimmerman, 1990). The next three sections analyse the participants' goal setting behaviour as it relates to the planning, composing and reviewing processes, as well as their intentions to continue learning the following academic year.

6.4.1: Goal Setting Related to Planning, Composing, and Reviewing

After prompting participants to consider how they spent their time and the strategies they used when planning, the template prompted them to set goals to improve future planning for writing. Two types of participant chose not to set any planning goals: those who were happy with their planning and thus felt no changes were necessary, and those who did not respond to the prompt.

Table 40: Satisfaction with Planning and Subsequent Goals per NCEA Group

	n	Happy so no goals set	Happy but set goals	Not happy so set goals	Did not Respond
All	71	10%	55%	14%	21%
Excellence	26	15%	58%	4%	23%
Merit	30	7%	63%	13%	17%
Achieved	15	7%	33%	33%	27%

As indicated in Table 40, there were a few more Achieved participants who did not set any planning goals (27% cf. 23% and 17%) and a few more Excellence participants who were happy with their planning and set no new planning goals (15% cf. 7% and 7%). However, most participants in all three grade groups were in the category of those who were happy with their planning but could see ways to improve it, 58% of Excellence participants, 63% of Merit and 33% of Achieved.

An equal number of Achieved participants (33%) set goals for improved planning because they were not happy with what they had done previously, compared with 13% of Merit participants and 4% of Excellence. Overall, approximately two-thirds of participants indicated by their goals that

they were motivated to improve their planning process in future writing events (62% of Excellence participants, 76% of Merit, and 66% of Achieved).

As indicated in Table 41, the three most frequent goals set by participants in all three grade groups were spending more time, monitoring better and memorising/remembering more. The Excellence group had the smallest percentage of participants who planned to spend more time (15% compared with 37% and 33%) or memorise more (8% compared with 27% and 20%) while the Achieved group had the lowest percentage of participants who planned to monitor better (20% compared with 31% and 37%).

Table 41: Types of Planning Goal per NCEA Group

	n	Spending more time	Monitoring better	Memorising more
Excellence	26	15%	31%	8%
Merit	30	37%	37%	27%
Achieved	15	33%	20%	20%

Based on the groups' reported planning behaviour and the goals they set to improve future planning, the Excellence participants clearly set goals that reflected their previous planning behaviour; the Merit and Achieved participants set appropriate goals to commit more time, memorise/remember more and monitor better, although they could be encouraged to memorise language rather than scripts; and the Achieved participants, in particular, could be encouraged to spend more time critically monitoring their language use. The goals set, therefore, did not necessarily match reported lacks in planning. It seemed that Merit and Achieved participants, in particular, needed guidance to link planning goals with previous planning behaviour.

While the Term 1 and 2 Merit intervention groups set a similar number of goals overall, more Term 1 participants were involved in setting goals (43% compared with 31% wanted to spend more time and monitor better, and 36% compared with 19% wanted to memorise/remember more). The goal setting behaviour of the Term 1 Merit group compared with the Term 2 was in line with their lower achievement (see Section 6.3.1).

The reflection template also prompted participants to set goals related to future composing processes. Of the 71 participants, only 38% responded to the prompt, 39% of the Excellence group, 43% of the Achieved group and 33% of the Merit. However, most of the goals mentioned in the composition reflections related to planning rather than composing. Planning goals included remembering better (8.5%), knowing verbs better (7%), learning spelling and accents better

(5.6%), and committing more time to practising language in advance of the writing test (4.2%). Only one goal mentioned by three or more participants (7%), using more varied and complex language, related directly to the composition process. Other composition goals mentioned by one or two participants included better time management, focusing on what you know and keeping it simple, managing stress, and leaving space in hand-written scripts to add and edit language when monitoring.

After receiving feedback on their first drafts, the reflection template prompted participants to tally their most frequent errors, to record their reaction, and to set goals for writing their second draft. Examples of blog entries written by Sharon, an Achieved participant, Bao-Zhi, a Merit participant, and Richard, an Excellence participant, follow.

> My top 3 most common errors are my verbs, agreements and i need to check if my words are all in the same tense. But i'm not really surprised at that because i've always known that their my weakest ones especially verbs. But i know it's something i can work on and improve on. I think i need to revise them first then i can go back over my work. (Sharon, Achieved)

> My 3 most errors were agreements, spelling and verbs. I'm not surprised that spelling was my worst because I have always struggled with spelling in French. I am going to do what the annotation code sheet says and I will use the strategies that it recommended. I hope that learning my common errors will make me think more about the when I go to do my second draft. (Bao-Zhi, Merit)

> My most frequent errors were articles, particles and conjunctions (L), Idioms and formulaic expressions versus franglais (U), verbs (V) and omitted words (^). I'm not very surprised as I find verbs and articles most difficult when writing in french. I intend to practice and study verbs and tenses more and figure out some ways to remember how to use the correct articles in my writing. (Richard, Excellence)

As indicated in Table 42, the most mentioned goal for all three groups was to improve monitoring. Thirty-nine percent of participants said they needed to improve their second draft by monitoring better, 24% intended evaluating the gap and accessing resources to overcome the gap, 23%

intended reviewing language before rewriting, and 18% planned to practise language before rewriting. Thirty-two percent did not respond.

Table 42: Goals Set in Response to Feedback in Group Percentages

	n	Monitoring	Reviewing	Practising	Evaluating the Gap	No Response
All	**71**	39%	23%	18%	24%	32%
Excellence	**26**	38%	12%	15%	31%	35%
Merit	**30**	40%	27%	20%	27%	30%
Achieved	**15**	40%	33%	20%	13%	33%

When considering the goals in grade groups, there was a similar percentage in all three groups who intended monitoring (38%, 40%, 40%) and practising (15%, 20%, 20%), and a similar percentage who failed to respond (35%, 30%, 33%). However, their goals related to reviewing and evaluating and overcoming the gap echoed the pattern in the planning strategies. The more general and non-specific strategy of reviewing was mentioned by more Achieved participants (33%) and fewer Excellence participants (12%), and the more critically sensitive strategy of evaluating and overcoming the gap was mentioned by more Excellence participants (31%) but fewer Achieved (13%). For Merit participants, the same percentage (27%) mentioned reviewing as evaluating and overcoming the gap. Like their strategy use, the astuteness of the participants' goal setting behaviour in response to feedback seemed to approximate their L2 achievement levels.

6.4.2: Goals to Continue Learning

After considering goals for improved writing, the study researched the participants' intentions regarding French for the following academic year, Year 12. For this analysis, their L2 achievement data were combined with data from their personal profiles. In the profile surveys which were conducted towards the end of the academic year in which the research took place, the participants recorded their final internally assessed NCEA portfolio grade and their intentions for the following year. Because five of the 71 participants had left their respective schools between the completion of the research at the end of Term 2 (mid-year) and the NCEA submission date at the end of Term 3, only 66 of the 71 research participants completed the survey. Only these 66 respondents are considered in this section.

To begin with, the NCEA indicator grades awarded to the participants' Task 1 Draft 1 in Term 1 were compared with their final portfolio grades awarded at the end of Term 3. It was interesting to

note that 37 (56%) participants received the same grade for their NCEA portfolio as they did for their first assessment draft of the year (see Figure 12). Even so, the NCEA grades for 19 (29%) participants went up in relation to the first draft while the grades for 10 (15%) participants went down. There was no apparent pattern in this movement apart from trending more up than down, as might be hoped. Ultimately, six Achieved grades became Merits, thirteen Merits became Excellences, three Excellences became Merits, six Merits became Achieveds, and one Achieved became a Not Achieved.

Figure 12: Task 1 Draft 1 and Final NCEA Grades with Retention Rates
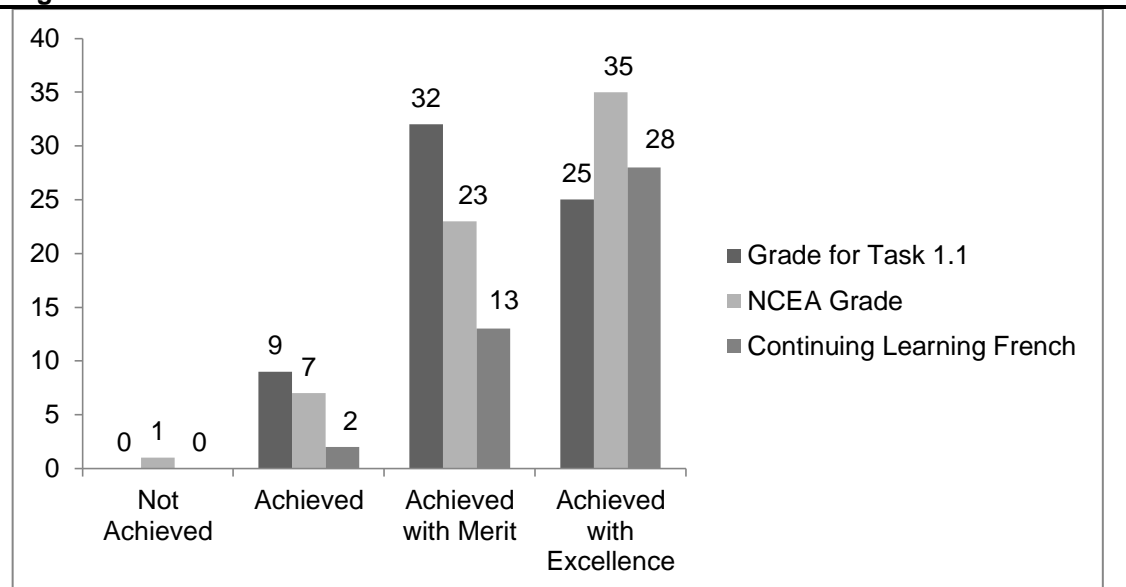
Secondly, the analysis compared the participants' final NCEA grade and Task 1 Draft 1 grade with their goals for continuing learning. As indicated in Figure 12, the number of participants intending to continue learning increased as the NCEA grades increased; the extrinsic motivation of the final grade appeared to influence retention. Of the 66 participants, 43 (65%) indicated their intention to continue learning, two (29%) Achieved participants, thirteen (57%) Merit and 28 (80%) Excellence.

Thirdly, the analysis investigated the movement of first draft grades between Task 1 and Task 2 with goals to continue learning. Of the 23 participants who chose not to continue, one (5%) advanced a grade between Task 1 and 2, four (17%) dropped a grade, and 18 (78%) retained the same grade. Of the 43 who chose to continue, six (14%) advanced a grade, nine (21%) dropped a grade, and 28 (65%) retained the same grade. The awarding of grades to drafts did not appear to parallel retention and attrition whereas the final portfolio grade did. Participants were clearly

aware that draft grades represented work in progress and, as such, were able to be improved (see Section 6.6.1). This seemed to indicate a certain level of self-efficacy.

This section examined the L1 blogs according to the Motive dimension of self-regulated learning. It was found that, while two-thirds of participants were happy with their planning behaviour, two-thirds also set goals to improve their planning; and while the goals of the Excellence participants seemed appropriate to their reported planning behaviour, the goals of the Merit and Achieved participants did not necessarily match their planning needs. In the analysis of the participants' intentions to continue learning or not, it seemed that the higher the final internally assessed NCEA portfolio grade, the more likely the participant was to continue language learning the following year. However, where the NCEA grade appeared to be a high extrinsic motivator, the first draft grades did not appear to affect intentions for learning the following year. Participants seemed to understand that they were able to improve these grades in later drafts. What was most interesting was that more than half the participants received the same grade for the first draft of their first writing assessment in Term 1 as they did for their final internally assessed portfolio at the end of the academic year. Therefore, to track daily learning, other measures as well as the NCEA rubric are clearly needed.

6.5: The Physical and Social Environment Dimensions

Having considered the self-regulation dimensions of Time, Motive, and Method, the Physical Environment and Social Environment Dimensions are now considered. The Physical Environment dimension refers to the places where self-regulated learners choose to learn and how they restructure their environment to maximise their learning. The Social Environment dimension refers to the social sensitivity and resourcefulness of the self-regulated learner who uses social networking and selectively seeks help to move their learning forward (Schunk & Usher, 2013; Zimmerman, 1990).

6.5.1: Physical Environment Dimension

In order to self-regulate, learners need to be able to make choices (Schunk & Usher, 2013). The physical environment of the classroom test, the initial feedback and editing, and the homework of the planning and rewriting processes were aspects that were determined for the participants in this study and on which they were not asked to reflect. However, several referred to relocating their books from school to home when planning and several mentioned the places

where they planned. For example, Chrissy, an Achieved participant, wrote: "I am quite happy with the preparation but next time I need to remember to bring more books from school." Nancy, a Merit participant, recorded that she studied "at home, in class and with friends," and Steve, an Achieved participant, went "to the library" to study with friends.

6.5.2: Social Environment Dimension

Where the physical environment was a minor factor in the research, the social environment was a foundational factor because of the interaction that occurred around feedback. This is considered in detail in the next section. Apart from the interaction that occurred around feedback, only four participants referred to other social aspects of learning. Nancy and Steve, in the previous section, indicated working with others as they prepared for the writing test and two others mentioned seeking help from their teacher. Isla, an Excellence participant, wrote regarding her planning: "I also asked (the teacher) for help with any of the sentences that I wasn't sure about." Ethan, a Merit participant, wrote regarding his reviewing: "I intend to study more and ask the teacher more questions before I write my 2nd draft." It appears, however, that most participants planned and reviewed their writing alone.

This section considered the L1 blogs according to the Physical Environment and the Social Environment dimensions of self-regulated learning. In order to self-regulate, learners need to be able to make choices. For much of the research period, the choices related to physical and social environment were made for the participants by the task design and the context. All participants wrote their tests and responded initially to feedback in class and most planned and reviewed their writing at home. However, several participants mentioned transferring resources from school to home and several mentioned working with others or seeking help from the teacher.

6.6: The Behaviour Dimension

The sixth and final dimension of self-regulation to consider in the L1 blogs is the Behaviour Dimension. This refers to the actions that learners take to monitor and evaluate their progress towards their goals. As they work on tasks and observe and judge their progress, self-regulated learners accept responsibility for their learning outcomes; they do not attribute their progress or lack of progress to causes outside their control. If learning is going well, they react by continuing with what they are doing; if learning is not going well they change direction; and as they progress toward their goal, their perseverance and self-efficacy increase (Macaro, et al., 2016; Schunk &

Usher, 2013; Zimmerman, 1990). In this section, the participants' evaluation of progress when planning, composing and reviewing their writing is analysed, as is their evaluation of the entire writing process.

6.6.1: Evaluating Planning, Composing and Reviewing

The reflection template prompted the participants to consider whether they were happy with the time they had spent and the strategies they had used when planning in the seven days prior to the test. As a result of their response, they were prompted to set goals for improving planning processes in the future. While 21% of participants did not respond to this prompt, 65% reported that they were happy with their planning; 10% were very happy and therefore chose not to make any changes and 55% were happy but could still see ways to improve. The remaining 14% were not happy, and thus set goals to improve. Of the participants who indicated that they were very happy or largely happy with their preparation, 40% were Achieved, 70% Merit and 73% Excellence.

Their self-observation and self-judgment as evidenced in their satisfaction or lack of satisfaction with their planning behaviour seemed to indicate balanced and astute responses; the lower the NCEA grade awarded the first draft, the more dissatisfaction was expressed. Where 33% of Achieved participants were unhappy, and consequently set goals, this was the case for 13% of Merit participants and 4% of Excellence participants (see Section 6.4.1). In order to improve, learners need to be accurate in their self-perceptions (Macaro, et al., 2016). However, while the satisfaction/dissatisfaction recorded and the number of goals set seemed appropriate, the content of the goals did not always match prior behaviour. Participants needed more guidance in this area (see Section 6.4.1).

As well as referring to 'happiness' about their planning as they were prompted to do, six participants indicated self-observation and self-judgment by referring to other emotions. Three referred to confidence or strategies to gain confidence. Maria, a Merit participant, said: "I was confident in my studies and believed that I would do well." James, a Merit participant, wrote: "I think for next time I would use LanguagePerfect[14] more to prepare, as well as prepare more in

---

[14] Language Perfect is a popular online language learning programme that was developed for use in New Zealand language classes but is now used in many subject areas and by schools in many nations around the world. The Language Perfect team have consequently changed the programme's name to Education Perfect.

advance, so that I would be more confident." Eva, an Achieved participant, added: "I would definitely spend more nights preparing for this and gaining more confidence in my writing."

The other three referred to nervousness when composing. Olivia, an Excellence student, stated: "I probably should have remembered better, although my forgetfulness was probably based on nerves." Jonathan, a Merit participant, wrote: "I was nervous going into the test though. Next time I want to learn more useful phrases and have some sentences already created in my mind." William, an Achieved participant, added:

> I think once I got under test conditions I put so much pressure on myself that I forgot a lot of simple Level 1 tenses. During the test I was so anxious and worried about failing that I couldn't wrap my head around simple French, my brain automatically went to the english way of speaking and not the French state of mind.

Of the six participants who mentioned a lack of confidence or nervousness related to planning and test-taking, two did not appear to manage those feelings by 'activating supportive emotions, beliefs and attitudes' (Oxford, 2011). One strategy that seemed to help other participants was reasoning that the test was just the first step in the process writing unit. Natalie, an Excellence participant, wrote: "minor mistakes in the first draft don't concern me at all, and most of my mistakes were minor and I will be able to correct them using the resources suggested." Khan, a Merit participant, wrote: "even if I'm wrong I know I can fix it for the second draft." (See also Section 6.4.1)

In addition to evaluating their planning process, participants were prompted to evaluate their composing process (see also Section 6.3.2). While 97% of participants mentioned composition difficulties and 56% explained how they overcame those difficulties, only 35% articulated successes, 42% of Excellence participants, 33% of Achieved and 33% of Merit.

Table 43: Successes when Composing in Group Percentages

	n	Remembering Content	Writing Well	Remembering Language	Good Time Management	Not Mentioned
All	71	16%	9%	6%	4%	65%
Excellence	26	12%	12%	0%	8%	58%
Merit	30	17%	10%	10%	3%	67%
Achieved	15	20%	7%	0%	0%	73%

However, as was the case with composition goals, most of the so-called composition successes related more to planning than composing (see Section 6.4.1); successes mentioned by three or more participants (see Table 43) included remembering content (16%), remembering language (6%), non-specific comments about 'writing well' (9%), and managing time during the test (4%). When comparing the successes in NCEA groups, the most mentioned by all groups was remembering content. However, 10% of Merit participants also mentioned remembering language. This is encouraging as it balances with the 11% of Merit participants who mentioned rote learning of scripts (see Section 6.3.1). From the analysis, it seems that participants were wary of articulating what went well when composing, and when they did engage, they tended to focus on successful recall during the test.

In addition to evaluating their planning and composing processes, participants were prompted to evaluate their review process. In the feedback lesson when their first drafts were returned, the template prompted them to tally their error codes and to focus on their two or three most frequent error types only. Almost two-thirds of the participants (62%) indicated that they were not surprised by their tallies of most frequent error types. This response revealed a level of accurate self-judgment conducive to improving learning. However, 15% were surprised by all their frequent error types and 10% were surprised at one or two error types but not all. Thirteen percent did not respond to the prompt.

The reactions of the three grade groups were comparable. A similar percentage of each group was not surprised at their most frequently committed errors (66%, 60%, 60%), a similar percentage was surprised (15%, 13%, 20%), a similar percentage had a mixed response (8%, 13%, 7%), and a similar percentage did not respond (14%, 13%, 12%).

Table 44: Percentage of Groups Finding Feedback Types Helpful

	n	Indirect Error Codes	Teaching on Common Errors	Feed Forward Comments	NCEA Grade Indicator
All	71	62%	44%	44%	44%
Excellence	26	54%	42%	27%	42%
Merit	30	70%	37%	43%	33%
Achieved	15	60%	60%	73%	33%

After evaluating their planning, composing and reviewing processes, the template prompted participants to evaluate the helpfulness of the feedback for improving their second draft. This was another attempt to encourage critical awareness. As indicated in Table 44, almost two-thirds of

participants (62%) found the indirect error codes to be the most helpful, and 44% found the other three forms helpful. Where 62% said the indirect codes helped them quickly identify their errors and improve their writing (54% of Excellence participants, 70% of Merit and 60% of Achieved), 19% had mixed feelings and 19% did not respond. The 19% who had misgivings said they could not spot the error or, having spotted it, they did not know how to fix it. Comments from four Merit participants follow:

> The codes were kind of helpful at a really base level because all they really did was point out I had made a mistake somewhere in the sentence and not really much more than that ... sometimes I'd make a verb error and there'd be more than one verb and it's like how am I supposed to know. I guess that is the point. (Dewei)

Jacqueline wrote along the same lines: "The codes are helpful, showing where I really need to work on. The only way it would be better is if they were fixed for me. Of course that would be cheating and actually helpful." Peter also found the codes frustrating. He said: "I'm worried I will correct my first draft incorrectly and make more errors in my second draft." Jessica added: "I just wish it was more straightforward and the teacher could tell me exactly what was wrong."

Regarding the teaching on common errors, 44% of participants found this form of feedback helpful, 7% found it sometimes helpful, 10% found it not helpful, and 39% did not respond. When comparing responses in NCEA groups, more Achieved participants (60%) than Merit (37%) or Excellence (42%) found the teaching helpful because: their errors were similar to those made by others in the class; the teaching was explicit and cautionary; and they liked the social interaction. Tamati, an Excellence participant, stated: "The teaching on common errors was very useful too as there were some things i didn't know how to fix on my own and obviously a lot of others didn't know either and it was nice getting feedback verbally." Lauren, also an Excellence participant, wrote: "The common errors informed of errors I could have made, which would be useful in the future, but did not pertain to the errors in my writing." Rachel, an Achieved participant, wrote: "The teaching on common errors reminded me of stuff I already knew but forgot for the test." Celeste, another Excellence participant, added: "Teaching on common errors is good because it helps the whole class."

Regarding the feed forward comments, 44% of participants found this form of feedback helpful, 73% of Achieved, 43% of Merit and 27% of Excellence. Included with the comment was an NCEA

grade indicator. Overall, 44% of participants found the indicator useful, 42% of Excellence participants, and 33% of Merit and Achieved. Lan-Lan, an Achieved participant, Tanya, a Merit participant, and Selina and Maria, Excellence participants, wrote:

> They are helpful comments as I know where I am and how much I need to improve to pass. I understand that French is certainly not my strongest subject, and is hardest for me to learn and understand but the indications let me know what I need to improve on. (Lan-Lan, Achieved)

> The comments left by the teacher are very helpful as they go into more detail on where I stand and what I need to do to reach the next grade. They have gone into more detail as to where I have gone wrong and where I have done well. (Tanya, Merit)

> The teacher comment was motivating as she said 'just E' so that makes me want to make it so it isn't 'just E'. (Selina, Excellence)

> The comments and grade indicator helped as the comments reiterated the areas I need to work on and the grade indicator told me that I need to fix errors and perhaps add to my writing to get it up to a solid E level. It is helpful to know what my strengths and weaknesses are and how those correspond to my grade level. (Maria, Excellence)

However, 6% of participants found the comments and grade indicator unhelpful. Yuki, a Merit participant, wrote: "I was sort of disappointed when I got my suggestive grade, but I know what I need to do to improve next time. I'm not very sure whether or not redrafting my original will improve my grade." Patrick, an Excellence participant, wrote: "They are not too helpful as they didn't give me anything to work on, but the coding was helpful." Riya, an Achieved participant, wrote: "They helped me understand the level I was writing at, but I wish they wrote what else I could have done to improve it." Thirty-one percent of participants did not mention the teacher comment and 52% did not mention the grade indicator.

As well as indicating the helpfulness of the various forms of feedback, several participants went beyond the prompts to critique the task.

The teaching on common errors was relatively helpful, however, it would be nice to incorporate a code with the error. For example, if the error is using 'en' instead of 'a', or vice versa, indicating that if you have an 'L' code. (Keith, Excellence)

I really liked the codes. The explanations really help to find the errors way faster. That is really good. I would like it even more, if there would be examples next to [the codes in the schedule], but the way it is now I am happy with. (Rebecca, Merit)

This constructive criticism indicated the participants' depth of thinking. Their suggestions will help improve the feedback process that was used in this study and that is currently used by teachers who contributed to its creation and distribution via the NZAFT listserv.

6.6.2: Evaluating Final Comments

Having responded to the prompts in the reflection template, participants were offered an opportunity to add anything else they wished. More than half (54%) did not respond. Of those that did, most summarised the goals they had set in previous sections of the blog as if to remind themselves of their intentions, but some added a summarising or global statement. These two types of comment seemed to correspond with two of the cognitive strategies listed by Oxford (2011): conceptualising with details, and conceptualising broadly. Where 50% of the Excellence group conceptualised with details by reiterating their goals, 12% conceptualised broadly by making a summarising statement; 47% of the Achieved group reiterated their goals while 13% summarised the process; and 40% of the Merit group reiterated their goals while 7% summarised the process. The reiteration of goals indicated a high level of motivation, perseverance, and self-efficacy as well as extended abstract thinking. The global statements also indicated extended abstract thinking (Biggs & Collis, 1982).

The participants who conceptualised with details often listed what they wanted to remember. For example, Shiloh, an Excellence participant, wrote: "Remember to proofread and watch out for verbs." Michael, a Merit participant, promised himself: "I will do my best to think about my second draft thoroughly, and do my best to stop making grammatical mistakes by reading over each sentence carefully." Annie, an Achieved participant, listed the resources she intended using in subsequent drafts:

> Language Perfect - the sentence structures and verb usage. Language Perfect - the word remembrances (NCEA List). Year 10 French Revision. Studio Trois Exercise Book. Cahier Gris Grammar Structures Book. Practising writing at home, using Studio Trois Livre to help.

The participants who conceptualised broadly wrote summarising statements on the portfolio process itself or offered thanks or a touch of humour. Comments on the portfolio included:

> The techniques were good and easy to follow so hopefully that would help me in my writing portfolio. (Miriam, Achieved)

> I feel like this whole process was very rushed, and I would've preferred to have our books rather than a week to prepare. (Rowena, Achieved)

> I find that the draft gets easier as the year goes on, fewer mistakes are made and more language is used. (David, Merit)

> I like the format of the 30 minute written test, then fixing errors, then taking my work home to fix. It has worked very well for me (Debbie, Excellence).

Comments of thanks and humour included:

> i love how two teachers will grade this as i will get more feedback to improve my grades each time and gain more knowledge in my french :). (Hayden, Merit)

> Thank you for correcting all of my errors and making my writing better. (Sarah, Excellence)

> Please excuse my French :). (Jacob, Excellence)

These latter entries are a reminder that, although the reflective blogs were designed to increase language learning through self-recording, self-observation, self-judgment and self-reaction, they also offered an opportunity for intentionality-reciprocity (Feuerstein & Feuerstein, 1999) and provided a platform for the dialogic engagement between student and teacher described by Slavkov (2015).

In this section the L1 blogs were analysed according to the behaviour dimension of self-regulated learning. From an analysis of the reflections related to planning, it seems that two-thirds of participants were happy with their planning process, but the higher the NCEA grade, the more

likely they were to express satisfaction. Even so, one third of participants from all three NCEA groups reported that they regretted not spending enough time relative to the demands of the writing test. From an analysis of the reflections related to composing, it seemed that, the higher the NCEA grade, the more able and willing participants were to indicate their composition successes. From an analysis of the reviewing process, it seemed that almost two-thirds of the participants were not surprised at their most frequent error types which indicated astute self-observation and self-judgment; and when evaluating the four types of feedback they received, almost two-thirds of participants (62%) found the indirect error codes to be helpful for improving second drafts, and 44% found the other three types of feedback, teaching on common errors, feed forward comment and NCEA indicator grade, helpful for improving second drafts.

6.7: Summarising the Chapter

This chapter has presented a broad range of findings from an analysis of the L1 blogs written by the Year 11 participants who took part in the reflection intervention, either in Term 1 (n = 27) or in Term 2 (n = 44). It sought insights regarding the factors that might have impacted on the amounts and types of L2 gain that occurred after L1 written reflection, it sought insights into the participants' self-regulation, and it sought any further insights that might enlighten L2 teaching and learning. To this end, after analysing the language in the blogs for linguistic complexity and evidence of depth of thinking, the blogs were investigated according to Zimmerman's six dimensions of self-regulated learning (Schunk & Usher, 2013) and the three writing processes, planning, composing and reviewing (Flower & Hayes, 1981; Macaro, et al., 2016; Schoonen, et al., 2009), and they were coded according to Graham and Macaro's (2007) formulation strategies, Oxford's (2011) language learning strategies, and Zimmerman's self-regulation strategies (1990). Key findings are summarised in Figure 13.

In the introductory section, the analysis of the language used in the blogs indicated that it was relatively simple and unsophisticated. It also indicated that, the higher the NCEA grade awarded the first L2 draft, the higher was the word count in the blogs, the more complex the lexis, and the deeper the thinking. Generally speaking, however, the blogs were not long or complex and they indicated a large amount of level 3 multistructural thinking according to the SOLO Taxonomy (Biggs & Collis, 1982).

As indicated in Figure 13, there were four important findings raised by the analysis related to the planning process. Firstly, it seemed that the higher the NCEA grade awarded to the first draft on which the participant reflected, the more time the participant was likely to spend on planning and the greater was likely to be the satisfaction with the planning. However, approximately a third of all participants (27% of Excellence, 33% of Merit and 40% of Achieved) regretted not spending enough time relative to task demand.

Figure 13: Chapter Six Findings

Planning	Composing
The higher the NCEA grade, the more... • time was spent • memorising occurred • effective the strategies used • strategies were used • strategies were used in combination • likely relevant goals were set	• Almost all mentioned difficulties • Half used strategies to overcome difficulties • One third articulated successes • One third set goals to improve composing but most goals related to planning not composing
Reviewing	**Continuing**
• Two thirds were already aware of their most frequent error types • Two thirds found the indirect error codes helpful • Four in ten found the other forms of feedback helpful	• Half received the same grade at the start and end of year • The final NCEA grade was associated with continued learning • Draft grades fostered self-efficacy

Secondly, the analysis indicated that the more time spent planning, the more memorisation of content, language or scripts was used during the planning period. Thirdly, the higher the NCEA grade, the more effective were the planning strategies used; the Excellence participants were more likely to use strategies that indicated critical sensitivity to language (monitoring, and evaluating and overcoming the gap). Fourthly, while two-thirds of participants were happy with their planning process, two-thirds also set goals to improve their planning; and while the goals of the Excellence participants seemed appropriate to their reported planning behaviour, the goals of the Merit and Achieved participants did not necessarily match their planning needs.

In the analysis of the reflections related to the composition process (which occurred under test conditions and without resources) several important points were raised (see Figure 13). While all but two participants recorded difficulties and just over half recorded strategies to overcome difficulties, only a third recorded successes; and when it came to setting goals for improving the composition process, most of those goals related to the planning process.

From the analysis of the reviewing process, important points were added. Firstly, almost two-thirds of the participants revealed a level of accurate self-judgment in their awareness of their most frequent error types. However, 15% were surprised at their error types and 10% were surprised at one or two of them but not all. By reflecting on their error types, these participants had the opportunity to improve their self-observation and self-judgment capabilities. Secondly, more participants indicated in their reflections that, of the four types of feedback offered, the indirect error codes were helpful for improving their subsequent drafts. The two-thirds who indicated this liked the thoroughness of the feedback and the way the codes highlighted errors they might not otherwise have noticed. Thirdly, the participants found the three other types of feedback equally useful. Just over a third found the teaching helpful because their errors were common to others in the class, the teaching was explicit and they enjoyed the face-to-face component; just over a third found the feed forward comments helpful because they indicated how to improve against the standard; and forty percent found the NCEA grade helpful because it indicated where they stood against the standard at that point in time which motivated them to improve.

In the analysis of the retention and attrition rates, further points were raised. Firstly, more than half the participants received the same grade for the first draft of their first writing assessment in Term 1 as they did for their final internally assessed portfolio at the end of the year. To provide extrinsic motivation and more fine-tuned assessment for teaching and learning, other forms of assessment are required. Secondly, the higher the final NCEA grade, the more likely the participant was to continue language learning the following year. Thirdly, where the NCEA grade appeared to be a high extrinsic motivator, the first draft grades were not. Participants understood that they were able of improve these grades in later drafts. This seemed to indicate a good level of self-efficacy.

In the final section of the blogs, some participants chose to remind themselves of their goals, some summarised or critiqued the writing process and some indicated their gratefulness for the help they received. These entries were a reminder that the reflective blogs offered opportunities for learners to share their thoughts with their teachers, for intentionality-reciprocity (Feuerstein & Feuerstein, 1999), and for dialogic engagement (Slavkov, 2015).

Having completed the quantitative analysis of the L2 data in the previous chapter and the largely

qualitative analysis of the L1 reflective blogs in this chapter, the next chapter revisits the research

questions in light of these quantitative results and qualitative findings.

CHAPTER SEVEN: DISCUSSING AND SYNTHESISING

The goal of this seventh chapter is to synthesise the results of the data analyses reported in the previous two chapters and to address the research questions posed in Chapter One:

1. Does L1 written reflection increase L2 learning when adolescents write for portfolio assessment? If so, what is the type and amount of increase and for whom?
2. What insights, if any, does the content of the L2 reflections reveal concerning the results of the L2 data analysis?
3. Are there any links between the reflective self-regulatory practices evidenced in the blogs, achievement levels, and/or continued learning?
4. What further insights regarding language learning do the blogs offer?

I addressed these questions via a mixed method, counter-balanced, comparative study that investigated whether the inclusion of L1 reflective practices in L2 portfolio assessment programmes increases opportunities for L2 learning, as demonstrated through L2 writing, and whether it is worthwhile to make time for reflective practices in busy secondary school language programmes. Appendix L contains examples of the L2 writing and matching L1 reflections of an Achieved, Merit and Excellence participant.

After conducting a pilot study in my own classroom to find suitable L2 learning measures, the current study took place in the FFL classrooms taught by four colleagues in four co-educational secondary schools different from my own. The Year 11 participants, 71 in all, were commencing their third year of French and their first of three years of the national qualification, the NCEA. The data for the research came from the first and second drafts of two tasks (therefore four drafts per participant) that were written during two four-week process writing units held three months apart, one at the end of Term 1 and the other at the end of Term 2 (mid-year). The intervention was a two-phase guided reflection that took place during the writing units and that was scaffolded by an online template containing five sets of prompts. The counter-balanced research design meant that the intervention group in the first writing unit became the comparison group in the second writing unit, and vice versa, which ensured that no student was advantaged or disadvantaged in this high-stakes NCEA assessment context. This also controlled for a possible order effect.

7.1: Interpreting the Results of the L2 Analysis

The potential impact of the reflection intervention (in terms of the type and amount of L2 learning gain and for whom) was measured in four investigations, each focusing on one of four types of language learning: 1. L2 written proficiency; 2. complexity, accuracy and fluency; 3. verb error elimination; and 4. self-regulated verb knowledge. The participants were placed into three pre-test groups according to the NCEA indicator grades awarded on the first drafts produced in each writing unit – Achieved (A), Achieved with Merit (M), or Achieved with Excellence (E). This produced different pre-test groups for each writing unit. Creating these groups enabled a more fine-grained comparison between intervention and comparison groups. The investigations compared the changes made to the L2 second drafts by the groups to establish any L2 learning gains and thereby the potential impact of the intervention.

7.1.1: L2 Written Proficiency

The first investigation used an analytic rubric developed by East (2008) to measure the impact of the reflection intervention on the core construct, L2 written proficiency. The analysis indicated that there was a statistically significant gain in the intervention group's proficiency in Task 1 but not in Task 2. When the Task 1 and 2 intervention groups were combined and compared with the Task 1 and 2 comparison groups, the analysis indicated that there was, once again, a statistically significant gain made by the intervention groups. When the intervention and comparison groups were broken down into intervention and comparison NCEA subgroups (A, M and E), the analysis revealed that the impact of the reflection intervention was statistically significant for the Task 1 Merit subgroup but not for the other two subgroups either in Task 1 or in Task 2.

It is important, therefore, to consider why a gain in proficiency for the Merit group, and not for the other two groups, was in evidence. From a sociocultural perspective, it seemed that the Task 1 Merit subgroup equated to Budoff's (1968) 'gainers' and to Sasaki's (2000) 'novices before instruction'. Because of the time in the academic year and because they were operating in their ZPD (Lantolf & Beckett, 2009), it seemed that these participants were ready for next steps to be taken in their interlanguage development (Williams & Burden, 1997). The reflection intervention seemed to provide them with the scaffolding they needed to engage in languaging and private

speech which research has shown to be instrumental in increasing L2 learning (Gal'perin, 1992; Knouzi, et al., 2010; Lantolf & Poehner, 2014; Polio, 2012).

The Achieved participants who did not seem to benefit from the reflection intervention could be categorised as Budoff's (1968) 'non-gainers' or Sasaki's (2000) 'novices after instruction'; it may be speculated that they needed more explicit support, in addition to dedicated time for reflection, before being able to progress. The Excellence participants could be categorised as Budoff's 'high scorers' or Sasaki's 'experts'; they had high grades in the first draft, were more self-regulated, and therefore, it would seem, had little need of dedicated time to reflect, their reflective practices appearing to be already routinised.

The simple act of recording reflections in writing also seemed to be important. It created a time and an occasion for thinking, for the kind of self-observation, self-judgment, and self-reaction that allows learners to track their progress over time (Schunk & Usher, 2013). As Schunk and Usher put it, "[s]elf-regulated learners observe their performance, judge their progress toward their goals, and react by continuing their approach or changing to a method they believe will be more effective" (p. 10). By offering the participants the time and guidance they needed to record their reflections, it seemed that the written reflection intervention facilitated these important self-regulatory behaviours and an increase in learning for the Term 1 Merit intervention group, in particular.

7.1.2: Complexity, Accuracy and Fluency

The second investigation measured the impact of the reflection intervention on complexity, accuracy and fluency in the participants' L2 writing. Where the East (2008) rubric was used in this study to calculate one overall score for L2 writing performance, the CAF analysis resulted in three separate scores. When considering complexity in the L2 writing, it was found that there was little to no increase between the participants' drafts, regardless of the NCEA group. In fact, based on the indirect coded feedback on error they received on their first drafts, some participants removed complex but error-laden sentences entirely, thereby reducing complexity rather than increasing it. There was a small amount (2-5%) of growth in complexity between the writing that occurred in Term 1 and the writing that occurred in Term 2. However, the Year 11 writing of all participants remained relatively unsophisticated at mid-year.

There are several possible reasons for this lack of change. Firstly, in this NCEA assessment context, it is the advised best practice for teachers to list possible content in English in the task instructions. If students have included this content in French in their first draft, there is less likelihood they would want to add more content in their second, especially as assessment conditions specify that feedback can only be given once (NZQA, 2016). Secondly, students and teachers must guarantee that the writing presented for assessment is the student's own work. First drafts are therefore often written in class under supervision. To maintain proof of authorship, students are aware that any changes they make to their second drafts must bear resemblance to their first drafts. This means they may be less inclined to add new content. Thirdly, feedback on error at this level of unsophisticated transactional language most often consists of word level feedback, so students respond by making mainly word-level changes. Although sentence-level feedback is also offered, students at this performance level (A2+ on the CEFR) find responding to indirect indications related to idiom and usage particularly difficult. They often therefore remove the offending sentences rather than running the risk of changing them wrongly. Fourthly, it is possible that some of the participants had not yet learned or had forgotten newly-learned connectors that would facilitate writing more complex sentences, such as, *bien sûr, cependant, ensuite, d'abord, d'habitude, heureusement* (of course, however, then, firstly, usually, fortunately), and so on. Allie, an Excellence participant, noted for example: "The difficulties I had were trying to link my sentences with suitable conjunctions and trying to make my sentences flow properly. I overcame it by using short sentences and opinions in my writing."

In addition to finding no statistically significant difference in complexity between intervention and comparison groups either in Task 1 or in Task 2, the investigation initially found that there were also no statistically significant differences in accuracy and fluency. However, when conducting a more fine-grained investigation of the NCEA subgroups, it appeared that there was a statistically significant difference in accuracy between the intervention and comparison Merit subgroups in Task 1 only and the effect size (p = .72) approached Cohen's (1988) convention for large effect (d = .80). This result aligned with the results of the first investigation.

Four observations may be made. Firstly, because accuracy is a component of the East rubric, it confirmed the reliability of the objective ratio of error free clauses per clause to measure accuracy and the reliability of the East rubric which required expert judgment to measure L2 written proficiency. Secondly, because the impact for the Merit participants occurred in Term 1 and not

Term 2, it aligned with Graham and Macaro (2007) who found that there was something to be gained from strategy instruction that begins early in the year despite the time it takes. Thirdly, because the late-elementary level participants were not inclined to add content to their second drafts, it raised the possibility that they were constrained by the conditions of the national assessment. They were not in a position to benefit from the reflection intervention in terms of complexity and fluency as fully as they might have been outside of the assessment conditions. Fourthly, it must be acknowledged, as Truscott (1996) claimed, that feedback on error, while it might increase accuracy, it might also inhibit learning gains in complexity as well as fluency. This certainly appeared to be case in this study. However, in their comparison of L1 and L2 writing strategies, Schoonen, et al. (2009) found that learners pay more attention to linguistic features when writing in a foreign language. They argued, "[t]hey seem forced to narrow their focus (more local readings) and become less fluent (smaller chunk size, fewer words per minute)" (p. 95). It may simply be that accuracy develops before complexity and fluency when learners are writing to learn language.

7.1.3: Verb Error Elimination

The third investigation was an error analysis which began by establishing the participants' most frequently committed error types using a schedule of very general error codes developed by teachers and approved by national NCEA assessment moderators. To increase self-regulation and a sense of agency in the feedback process, the participants were invited to tally and record their two or three most frequent error types based on the indirect error codes received and to select from a range of practice exercises linked from the assessment schedule to aid rule understanding and rewriting (Ellis & Shintani, 2014; I. Lee, 2007). The tally of error types revealed that spelling, nominal gender and number agreements, and verbs were the most frequent types of error made by the participants. The analysis also found that the number of participants with verb errors in their top three error types rose from 69% to 90% during the research period while the number with spelling errors and nominal agreement errors in their top three error types decreased during the research period from 86% to 69% and 72% to 58% respectively. The participants indicated their concern for verb error in their blog entries. Miriama, an Excellence participant, wrote:

My most frequent errors are verbs ... Because I think that verbs can be very difficult sometimes to correctly write in their appropriate tense and so I am not surprised that

it is my most frequent mistake. I believe I have corrected most of my verb errors but I will later check if I am correct so I don't make them any worse than they are at the moment.

There are three possible reasons for this increase in verb error and decrease in spelling and nominal agreement errors. Firstly, in the typical Year 11 French class, several new verb tenses are introduced in the first half of the year to facilitate the production of language that references past, present and future events (as required by NCEA Level 1 assessments). Secondly, nominal agreements and the principles of the French sound-spelling system have been taught in Years 9 and 10, the first two years of language learning, so, by Year 11, inaccuracies in these two categories are most likely to be, from a cognitive-interactionist perspective (Corder 1967; Ellis & Barkhuizen, 2005), mistakes rather than errors (because they result from slips rather than lack of knowledge) or, from a sociocultural perspective (Aljaafreh & Lantolf, 1994, Knouzi, et al., 2010), occurrences at a more intramental or self-regulated stage of L2 development (because the learner is able to assume responsibility for error correction without intervention from others). Thirdly, the participants were encouraged to use the newly learned verb tenses in the second writing task which they were not able to do when they wrote the first task. The change in task difficulty at mid-year may also have contributed to the increase in verb error. Regardless of the reason, as a result of the increase in verb error and the importance of verb knowledge in the Year 11 programme, the verb was chosen as the focus of the subsequent error analysis.

The error analysis sought evidence of the potential impact of the reflection intervention on verb error elimination. It was found that there was no statistically significant difference between the intervention and comparison groups, whether analysed intergroup or intragroup or in NCEA subgroups. Whereas the East (2008) and accuracy analyses resulted in statistically significant differences for the Task 1 Merit group, the verb error elimination analysis did not. This may be because different language features develop at different rates (Benevento & Storch, 2011; Torras & Celaya, 2001). Given that several new verb tenses are taught in the first half of Year 11, it is possible that all participants in the first half of the year required more explicit support, in combination with reflection, before more independent development in verb knowledge was possible. A verb microgenesis study was therefore conducted to further investigate this possibility.

7.1.4: Verb Development

The fourth investigation examined the participants' microgenetic verb development by combining data on correct verb production with data on revision outcomes and by measuring gain with a rubric developed from Aljaafreh and Lantolf's (1994) Regulatory Scale. It was found that the reflection intervention had no statistically significant impact on the microgenetic development of verb knowledge, whether analysed intergroup or intragroup or in NCEA subgroups. This supported the interpretation of the verb error elimination investigation that participants, regardless of NCEA grade group, required more explicit support, in addition to reflection, before their verb knowledge would improve.

7.1.5: Addressing Research Question 1

In answer to the first research question, therefore, the reflection intervention had a medium to large impact on the Term 1 Merit participants' L2 written proficiency and accuracy but had little impact on the complexity, fluency, verb error elimination or verb self-regulation as evidenced in their writing. It seemed that the Term 1 Merit group equated to Budoff's (1968) gainers and Sasaki's (2000) novices before instruction because they were able to take the next steps in their interlanguage development related to L2 written proficiency and accuracy but needed more support, in addition to reflection, before being able to develop other language features. It seems that the Achieved participants, equated to Budoff's 'non-gainers' or Sasaki's 'novices after instruction', who did not benefit from the reflection intervention, and who needed more explicit support in addition to dedicated time for reflection before being able to progress, regardless of language feature. The Excellence participants, who equated to Budoff's 'high scorers' or Sasaki's 'experts', had high grades in the first draft, were more self-regulated with regards to L2 written proficiency and accuracy, and therefore, it seems, had little need of dedicated time to reflect, their self-regulation of these features being already routinised and their language knowledge already internalised. However, given that different language features develop at different rates and times (Benevento & Storch, 2011; Torras & Celaya, 2001), it is possible that, with further support and continued opportunities to reflect, the other features of their written language - complexity, fluency, and verb knowledge - would also gain more than they would without the opportunity to reflect. Further insights and confirmation of the above interpretations were sought in the L1 reflective blogs.

7.2: Interpreting the Findings of the L1 Analysis

The reflection intervention involved the participants recording their self-observations, self-judgments and self-reactions in a two-stage L1 blog using shared Google Docs and scaffolded by a template containing five sets of prompts. Each participant took part in the reflection intervention either in Term 1 or in Term 2.

The analysis began by using Wordsift 2 (www.wordsift.org) to describe the language used in the L1 blogs and by using the SOLO Taxonomy (Biggs & Collis, 1982) to establish the depth of thinking evident in the blogs. It quickly became apparent that the language and the thinking were relatively simple and unsophisticated but that, the higher the NCEA grade awarded for the first L2 draft in the unit when the participant reflected, the higher the word count was in the blog, the richer and more complex the lexis, and the deeper the thinking. Generally speaking, however, none of the blogs were long or sophisticated and they showcased mainly multistructural level thought, that is, independent ideas that may be specific and detailed but that are not necessarily linked through cause and effect observations, comparisons, arguments, and so on (Biggs & Collis, 1982). This could possibly be due to the reflection being called a blog and participants using the type of language and thinking they might usually use when writing online. Regardless, it should not be surprising that the language in their L2 writing was also relatively simple and unsophisticated.

The next step in the analysis was to find insights in the content of the L1 blogs that might explain the amounts and types of L2 gain that occurred after the intervention, alongside insights regarding the participants' self-regulatory practices and achievement, and any further insights that might have implications for L2 teaching and learning. The investigation was guided by Zimmerman's (2013) six dimensions of self-regulated learning (Schunk & Usher, 2013), the writing processes of planning (conceptual preparation), composing (linguistic formulation and physical production) and reviewing (monitoring, editing and rewriting) reported in Flower and Hayes (1981), Macaro, et al. (2016), Paesani (2006), and Schoonen, et al. (2009), as well as Graham and Macaro's (2007) L2 formulation strategies, Oxford's (2011) language learning strategies, and Zimmerman's (1990) self-regulation strategies.

7.2.1: Planning Behaviour

Several insights were gained from the analysis of the participants' planning (and formulating) behaviour. They are discussed in this section according to 1. time, 2. goal setting, and 3. strategy use. Firstly, it seemed that the higher the NCEA grade awarded on the L2 first draft, the more likely it was that the participant had spent a larger amount of time on the first draft planning process. Even so, approximately a third of participants in all three groups regretted not spending enough time. In Zimmerman's self-regulation theory (Schunk & Usher, 2013), time is a *dimension* of self-regulated learning, and time management is a *process* of self-regulated learning but neither Zimmerman nor Graham and Macaro (2007), nor Oxford (2011) listed 'making time relative to task demand' as a *strategy*. Instead, *making time* is subsumed in other planning strategies, such as organising or researching. However, because so many of the adolescent participants in this research regretted the lack of time they spent, and because time spent approximated achievement level, it seems important to include 'making time' in strategy lists to increase attention paid to it by teachers and learners. It is interesting to note that, while one third of participants reported that they were not happy with time spent, two-thirds indicated that they were happy. Yet these two-thirds still set goals to improve their planning, which included goals to spend more time planning. This seems to indicate a high level of motivation and engagement in the portfolio assessment process and in the research, and an acceptance that increasing the amount of time spent in planning is likely to increase achievement levels.

The second set of planning insights arose from an analysis of the participants' planning goals. Although the template prompted the participants to set goals, the high level of response, two thirds of participants across all three grade groups, indicated a high level of motivation to improve. Their responses also indicated a high level of self-observation and self-judgment. Before knowing the indicator grade awarded to their first draft, many of the participants who gained the lower grades had already expressed dissatisfaction with their planning process, while many of the participants who had yet to learn of their higher grades had expressed satisfaction. Zimmerman (1990) argued that "[a]s children approach adolescence, their academic self-perceptions become more accurate" (p. 13). This is important to note because, in order for learners to become self-regulated and to learn language in a portfolio assessment context where only indirect feedback is permitted, learners need to be accurate in their self-perceptions (Macaro, et al., 2016).

While the participants' motivation might be high and their self-observation and self-judgment perceptive, their self-reaction was not as astute. Their satisfaction/dissatisfaction levels seemed to correspond with the number of goals they set, but the content of those goals did not always correspond well with lacks in their planning behaviour. The goals of the Excellence participants were more appropriate, but the goals of the Merit and Achieved participants could have been better aligned. Based on their planning and the grades they received, Merit and Achieved participants needed to set more goals related to memorising language rather than whole scripts, and more goals related to critically evaluating and overcoming their language gaps and monitoring their language. To connect goals with lacks in learning behaviour requires the type of relational thinking that is described in Level 4 of the SOLO Taxonomy (Biggs & Collis, 1982). Many participants indicated in their blogs that their thinking was at Level 3, multistructural, so it is not surprising that their ability to set effective planning goals for improving L2 writing is also at this level.

The third set of planning insights related to the participants' planning strategies. The analysis indicated that the more time the participants reported spending on planning, the more likely they were to report memorising or remembering content, language or scripts. Memorising is generally considered to be a strategy used by poorer language learners (Lamy & Hassan, 2003; Oxford, 2011), but, in this study, it was those participants who committed the most time to planning who memorised the most and it was these participants who gained the highest first draft grades. The strategy of memorising discrete language items or strings of items (also referred to as chunks or formulaic expressions) needs to be acknowledged because it ensures that the building blocks of language are set in place so that more complex understandings can later be constructed. These building blocks equate to the surface knowledge that Hattie and Yates (2013) claimed is essential but undervalued by educators who tend to promote deep knowledge to the detriment of the surface knowledge on which it has been built. If the SOLO Taxonomy (Biggs & Collis, 1982) is viewed as a learning sequence, then memorisation must equate to the Level 2 and 3 unistructural and multistructural stages where discrete facts or strings of facts are learned that will later be able to be connected (the relational level of the SOLO Taxonomy) and applied in new contexts (the extended abstract level of the SOLO Taxonomy). In foreign language classroom contexts, these lower or surface levels of thinking and language learning must not be avoided en route to the higher levels and en route to fluent language production. Even so, it must be acknowledged that,

in the current research, the memorisation of scripts rather than language by 11% of participants was not desirable. It is possible that initiating process writing units with a planning period followed by a test (designed to create benchmarks for the research and proof of authorship for NCEA purposes) moved the focus away from learning language onto learning for the test. If participants had been given less task information in advance of the test, it is possible that more language learning might have occurred. Indeed, another process writing sequence that omits tests altogether might be devised.

As well as providing insights into memorising, the planning strategy analysis indicated that by far the most frequently used strategy for all three groups was rehearsing by formulating practice drafts or sentences. The groups were also similar in their reporting of the strategy of activating prior knowledge, but only one in five from each group referred to it. This raises an important issue when analysing strategy use. Activating prior learning is integral to formulating practice sentences, so those in the four-fifths who wrote practice sentences but did not mention activating prior knowledge must not have been aware that they were using this strategy. Macaro (2001) explained that the knowledge of strategies, like L2 knowledge itself, moves from declarative to procedural through practice. Zimmerman (1990) found that, as learners become more self-regulated, their strategy use becomes more effective and their performance more routinised. Macaro, et al. (2016) referred to a dynamic continuum between linguistic knowledge and strategic knowledge. As linguistic knowledge becomes more fluent, less strategic knowledge is required, and as new knowledge is continually added, the balance continually changes.

A further insight into the participants' planning strategies was that, the higher the NCEA grade awarded for the first drafts, the more likely the participants were to use the more effective strategies. When planning by formulating practice sentences or drafts, the Excellence participants were more likely to evaluate their language gaps and take action to overcome them and then monitor the language they produced, whereas the Achieved participants were more likely to review language by studying notes, books and tests. Reviewing is a less effective planning strategy because it is non-specific and implies revisiting prior knowledge, whereas evaluating and overcoming the gap and monitoring language produced are more active and precise strategies and imply an awareness of specific lacks in knowledge and a desire to access resources or more expert language users who might help reduce those lacks. In the hierarchy of strategies

(Chesterfield & Chesterfield, 1985), evaluating and overcoming gaps in knowledge and monitoring language produced require more effort and skill than reviewing does.

Naiman, et al. (1978) described evaluating and overcoming the gap and monitoring language produced as "critical sensitivity to language use" (p. 5). Several early researchers found that use of strategies that indicate critical sensitivity to language distinguished between good and not so good language learners (Naiman, et al.; 1978; O'Malley, et al. 1985; Rubin, 1975; Stern, 1975). By comparing the three groups' use of the three strategies (reviewing, evaluating and overcoming the gap, and monitoring) it seemed that the Excellence and Merit participants were the 'better' language learners. Equally, it seemed that the choice of more effortful strategies might indicate motivation as well as achievement.

This line of thinking is also raised when considering the participants' orchestration of strategies. The current research found that the higher the NCEA grade, the more likely the participants were to orchestrate three or more strategies at a time. This aligns with Zimmerman and Martinez-Pons (1986) who found that higher-track mathematics students used more effective strategies, namely, researching information (which equates to evaluating and overcoming the gap), keeping records, monitoring, organising, and transforming, and they used more of these strategies in combination. Cohen and Macaro (2007) stated that "it is in the combination of strategies selected for a task and their orchestration through metacognition that sophistication lies" (p. 28). They found that ineffective language learners lack a range of strategies as well as the ability to use them in combination. They suggested three reasons for this: working memory limitations, lack of linguistic resources, and lack of motivation. When paired with the data on time spent, memorising, reviewing, evaluating and overcoming the gap, and monitoring, the comparative lack of strategy orchestration by the lower achieving participants could possibly indicate lower motivation in some participants in addition to lack of linguistic resources and working memory limitations. However, when adding goal setting behaviour alongside strategy use, motivation does not appear to be the main issue. Two thirds of Achieved students had set goals to improve their planning, and although less than a half of them set goals to improve their composing, theirs was the largest percentage of any group. Instead, it appears from their reflections that the lower achieving students did not lack in motivation but in the ability to select effective strategies and goals to match lacks in their learning, that is, they revealed less astute self-judgment and self-reaction skills than their counterparts.

Many researchers have found that, when strategies are taught and applied, achievement levels rise (Cohen & Macaro, 2007; Feuerstein, et al., 1991; Graham & Macaro, 2007; Grenfell, 2007; Guo, 2012; Oxford, 2011; Zimmerman, 1990). There is therefore hope for motivated lower achievers. This is because, in the L2 writing portfolio context, instruction on strategies and how to set goals has the potential to activate the cognitive functions that result in more effective language learning: noticing-triggering, hypothesis-testing and metalinguistic reflection (Fotos & Hinkel, 2007; Swain & Lapkin, 1995). Having interpreted the findings related to the planning process, the next section interprets the findings related to the composition process.

7.2.2: Composing Behaviour

In the analysis of the reflections related to the composition process (which occurred under test conditions and without resources) several more insights were provided. Firstly, only six participants mentioned having the time or needing the time to monitor their writing during the test. It is highly possible, therefore, that most participants spent the thirty minutes focused on formulating (converting ideas into language) and had little time or thought for monitoring. This could possibly indicate that these post-beginners found the formulation process difficult. Graham and Macaro (2007) indicated that British post-16 secondary school students of FFL also find the formulation process difficult. Roca de Larios, et al. (2008) found that less proficient L2 writers spend more time on formulating and less time on monitoring and evaluating their work than more proficient writers do.

Secondly, while all but two of the participants recorded difficulties when composing, just over half recorded strategies to overcome difficulties and only a third recorded successes, most of which related to recall. In addition, when it came to setting goals for improving composition processes, most goals related to the planning process instead; one third of goals related to remembering more. For the Achieved participants, two-thirds set goals to remember more. As indicated by Oxford (2011), those who engage in such surface strategies as rote learning produce poorer test grades than those who engage in deeper processing strategies, such as evaluating and overcoming the gap.

Thirdly, while it is understandable that more composition difficulties would be reported by the Achieved participants, it is interesting that more Achieved participants (67%) reported using strategies to overcome difficulties than did participants in the other two groups (61% of Excellence

participants and 47% of Merit). It seems logical that those who had the most difficulties would use the most strategies to overcome them, although this runs counter to the concept that strategy use approximates achievement level and could relate to, amongst other factors, motivation (Cohen & Macaro, 2007). Most Achieved participants in this study appeared to be highly motivated, so their lower achievement was more likely related to the other factors suggested by Cohen and Macaro, working memory limitations and lack of linguistic resources. From this study, it was also clear that their lower achievement related to lack of time spent in planning.

In general, from the analysis of the composition process, it appears that the participants did not understand it well. They appeared to need training to improve their ability to formulate by selecting and orchestrating effective strategies to overcome their difficulties, knowing what went well so it could be repeated, and setting goals that relate to lacks in prior behaviours. These skills represent the accurate self-observation, self-judgment and self-reaction that are necessary for self-regulated language learning (Macaro, et al., 2016). Having interpreted the findings related to both the planning and composing processes, the next section interprets the findings related to the reviewing process.

7.2.3: Reviewing Behaviour

When the first drafts were returned to participants with indirect feedback in the form of error codes in the margin of the line where an error occurred, the participants were invited to tally those codes to find their two or three most frequent error types. In addition, the error code schedule offered the participants links to online practice exercises related to each error type (see Appendix F). By discovering their errors themselves and choosing practice exercises to learn the language rules related to those errors, it was hoped that the participants would clarify the language features that they personally should focus on in order to improve their writing and, in the process, that they might develop a level of independence and agency. From the analysis of this reviewing process, several more insights emerged.

Firstly, almost two-thirds of the participants revealed that, prior to the tally, they were already aware of their most frequent error types; they were accurate in their self-observation and self-judgment. The reflection encouraged the 15% who expressed surprise at their error types and the 10% who expressed surprise at one or two error types, but not all, to self-observe and self-judge

more accurately. The honesty of this 25% in recording their surprise was a first step in achieving more accurate evaluation.

Secondly, the participants evaluated the helpfulness of the four types of feedback received. This evaluation was also designed to foster independence via self-observation, self-judgment and agency. Two-thirds of the participants indicated that the indirect error coding was helpful in improving their subsequent drafts because it was thorough and it highlighted errors they might not otherwise have noticed. Four in ten participants found the teaching on common errors helpful because their own errors were those common to the class, the teaching was explicit, and they enjoyed the face-to-face component; four in ten found the feed forward comments helpful because they indicated how to improve against the standard; and four in ten found the NCEA grade helpful because it indicated where they stood against the standard at that point of the year and motivated them to improve their second draft.

Thirdly, in response to the feedback received, participants were invited to set goals for improving their second draft. As indicated previously, very few participants seemed to have time to monitor their writing during the composition test. They were primarily focused on formulating, which they found difficult and with which they needed help. It was gratifying to note, therefore, that the goal chosen by 40% of Achieved and Merit participants and 38% of Excellence participants was to monitor their second draft better.

Apart from the desire to monitor their work better, the goals that the NCEA groups set in relation to reviewing and evaluating and overcoming the gap echoed the pattern seen in the planning strategies. The more general and non-specific strategy of reviewing was mentioned by more Achieved participants (33%) and fewer Excellence participants (12%), and the more critical strategy of evaluating and overcoming the gap was mentioned by more Excellence participants (31%) but fewer Achieved (13%). Apart from the desire to monitor their writing better, the astuteness of the participants' goal setting in response to feedback seemed to approximate their L2 achievement levels. Having analysed the findings related to the planning, composing and reviewing processes, the participants' final open-ended comments are now considered.

7.2.4: Open-ended Comments

In the final section of the blogs, the participants were prompted to reflect on anything further pertaining to their writing processes and strategies that they wished to add. Some participants

chose to remind themselves of their goals, some summarised or critiqued the writing process, and

some indicated their gratitude for the feedback received. These entries were a reminder that,

although the reflections were designed to increase language learning through self-recording, self-

observation, self-judgment and self-reaction, they also offered an opportunity for intentionality-

reciprocity, one of the three essential Feuersteinian mediations of learning (Feuerstein &

Feuerstein, 1999) and they provided an opportunity for the kind of dialogic engagement between

student and teacher described by Slavkov (2015). Through the dedicated occasions and the slow-

motion ponderings of the shared written talk, the participants were given the opportunity to

communicate safely in a way they might not otherwise have done without this opportunity to

express themselves. It seemed that the blogs, for some participants at least, were instrumental in

fostering the kind of learning relationships that increase the potential for language learning.

7.2.5: Motivation to Continue Learning

By adding the participants' end of year profile data to their achievement data, further insights

were gleaned. In the profiles the participants stated their intentions regarding whether they would

continue to learn French the following academic year or not. Three observations were able to be

made. Firstly, more than half the participants received the same grade for the first draft of their

Task 1 assessment in Term 1 as they did for their final internally assessed portfolio at the end of

the academic year. To provide extrinsic motivation as well as the more fine-tuned assessment

data necessary for differentiating teaching and learning practices, other forms of assessment, in

addition to the NCEA rubric, are therefore required. Secondly, it was apparent that, the higher the

internally assessed NCEA portfolio grade, the more likely the participant was to continue

language learning the following year. Yet, thirdly, where achievement success, as evidenced in

this final grade, appeared to be a high extrinsic motivator, the first draft grade indicators awarded

for the participants' writing throughout the year were not. It seemed that participants understood

that they were able of improve these grades in later drafts. This understanding indicated that the

portfolio assessment process was successful in creating a level of engagement and self-efficacy

that was conducive to language learning. So a strong connection between Assessment *for*

Learning principles as evidenced in writing portfolio assessment can be made.

7.2.6: Comparing Merit Groups

An important question to ask is why the Term 1 Merit intervention group gained significantly

more in L2 learning after participating in the reflection intervention than the Term 2 Merit

intervention group and, in fact, all other groups. A comparison of the L1 data of just these two groups indicated that, by every measure, the Term 1 Merit participants' behaviour resembled the Achieved groups' behaviour more, while the Term 2 Merit groups' behaviour resembled that of the Excellence groups more. The Term 1 Merit participants' blogs were shorter and showed a slightly lower level of thinking than the Term 2 Merit groups' blogs. A comparison of the types of planning strategies reported by each subgroup revealed that two-thirds (64%) of the Term 1 Merit participants mentioned the less skilful and less effortful strategy of reviewing compared with one fifth (19%) of the Term 2 group. All other strategies, including the more skilful and effortful strategies of monitoring and evaluating the gap were mentioned by more Term 2 participants than Term 1 (81% cf. 57% rehearsing, 63% cf. 14% evaluating and overcoming the gap, 38% cf. 21% monitoring, 38% cf. 14% memorising, and 19% cf. 14% activating prior knowledge). In addition to the differences in strategy use reported, there was a marked difference in the two groups' strategy orchestration. One fifth (22%) of the Term 1 group combined three or more planning strategies compared with one half (50%) of the Term 2 group. If strategy use is "strongly associated with superior academic functioning" (Zimmerman, 1990, p. 8), and L1 metalinguistic reflection is strongly associated with L2 proficiency (Alderson, et al., 1997; Renou, 2001; Simard, French & Fortier, 2007), then it seems that the Term 1 Merit group's strategy use indicated inferior achievement to the Term 2 groups' use.

Revisiting the L2 achievement data related to Task 1 Draft 1 confirmed that the Term 1 Merit intervention groups' L2 achievement sat lower on the achievement scale than the Term 2 groups' L2 achievement. Where the Term 1 group's L2 writing contained an average of 33.36 clauses, 4.5 dependent clauses and 11.57 error-free clauses, the Term 2 group's contained 38.95 clauses, 6.27 dependent clauses and 13.95 error-free clauses. Where the Term 1 group scored an average of 23.71 on East's (2008) L2 written proficiency scale, the Term 2 group scored 24.09. Where the first drafts of the Term 1 group contained an average of 31.04 correct verbs out of 37.85, the first drafts of the Term 2 group contained 32.05 out of 38.68. On every count the Term 1 Merit group scored lower in their first drafts than the Term 2 group. The Term 1 Merit group clearly had more scope for improvement both in L2 learning and in strategy use than their Term 2 counterparts. They therefore showed more potential for gain, and this appears to be evidenced by the significant differences noted earlier.

However, despite their poorer achievement and strategy use, the Term 1 Merit intervention group revealed astute self-observation and self-judgment. They indicated in their reflections that they were more dissatisfied with their planning behaviour than the Term 2 Merit intervention group and their goal setting behaviour rightly reflected this. While the Term 1 and 2 Merit intervention groups set a similar number of goals overall, more of the Term 1 participants were involved in goal setting – 43% compared with 31% wanted to spend more time, 43% compared with 31% wanted to monitor better, and 36% compared with 19% wanted to memorise more. It is therefore possible that the intervention which prompted reflection on planning strategies and then prompted goal setting relative to planning behaviour was a reason why the Term 1 Merit intervention students had a statistically significant increase in L2 written proficiency and in accuracy between drafts compared with their counterparts. It is also possible that setting aside time for reflection activated the supporting emotions, beliefs and attitudes necessary for the participants to boost their self-efficacy and persevere with improving their work by reducing the gap between the goals they set and the deficits of their first drafts.

7.2.7: Addressing Research Questions 2, 3 and 4

The analysis of the L1 reflective blogs aimed to answer the second, third and fourth research questions: 2. What insights, if any, does the content of the L1 reflections reveal concerning the results of the L2 data analysis? 3. Are there any links between the reflective self-regulatory practices reported in the reflections, achievement levels, and/or continued learning? and 4. What further insights regarding language learning do the reflections provide?

As stated above, the analysis of the L2 writing samples found that the Term 1 Merit intervention participants were the only group to gain significantly after the reflection intervention and they did so in L2 written proficiency and accuracy but not in the other features measured. By comparing this group's L2 writing and L1 reflection data with that of the Term 2 Merit group, it became apparent that both the L2 achievement and L1 reflection data of the Term 1 group indicated that they were at the lower end of the Merit band and had greater potential for improvement.

Interpreting these results from an interactionist-sociocultural perspective, the reflective blogs may be viewed as auto-input and private speech, and the reflection intervention as a mediation using the tools of dedicated time and the scaffolding of a template; and as participants were guided to

reflect on their writing processes, the cognitive processes conducive to language learning (noticing-triggering, hypothesis-testing and metalinguistic reflection) were activated.

Guided by Aljaafreh and Lantolf's (1994) regulatory scale to analyse the Term 1 Merit intervention group's L2 achievement and L1 reflection data, it seems that, in comparison with others, these participants were at a stage in their ZPD and in their self-regulation where they were able to benefit from the mediation of reflection time and reflection template and were able to take bigger steps in interlanguage development than would have been possible without the mediation. It seems that, by providing dedicated time and guidance for written talk, the reflection intervention encouraged the participants to engage in private speech or (auto-input), and, as such, activated the cognitive functions of noticing-triggering, hypothesis-testing and metalinguistic reflection which, in turn, increased their capacity to learn language (Knouzi, et al., 2010; Lantolf & Poehner, 2014; Smith, 2007; Swain & Lapkin, 1995).

If the Term 1 Merit participants, who sat closer to the Achieved group in achievement, benefitted from the reflection intervention in a statistically significant manner, why then did the Achieved participants not also enjoy such an increase in learning? Did the blogs offer any insights that would confirm the conclusion that the Term 1 Merit group equated to Budoff's (1968) 'gainers' and Sasaki's (2000) 'novices before instruction' because they were ready to learn and the mediation of the reflection intervention accelerated their progress? Were there any insights that the Achieved group equated to Budoff's 'non-gainers' or Sasaki's 'novices after instruction' because they needed more explicit support in addition to dedicated time for reflection before being able to progress? Was there any evidence that the Excellence group equated to Budoff's 'high scorers' or Sasaki's 'experts' because they had less need of dedicated time to reflect because they already had high grades and were already self-regulated learners?

Throughout the L1 blog analysis, the association between the achievement levels of the NCEA grade groups and self-regulation levels appears to be very strong. The Excellence participants, the highest L2 scorers, indicated in their L1 reflections that they wrote more and used a richer lexis and deeper thinking in their L1 reflections, that they were more likely to make more time for planning for the writing test, to remember more, to use more effortful and effective strategies (those that indicate critical sensitivity to language), to use these effortful and effective strategies in combination, to be more aware of and ready to express composition successes, to be more

aware of their typical error types, and to need less help linking learning goals to lacks in previous behaviours. They were closer in their L2 writing proficiency and accuracy and in their reflective behaviour to the self-regulatory end of the Aljaafreh and Lantolf's (1994) scale than the other two groups.

Equally, the Achieved participants, the lower L2 achievers, wrote less in their L1 reflections, used language that was not as rich and engaged in thinking that was not as deep, reported making less time to plan for the test and remembered less, used less effortful and less effective strategies, and fewer in combination, and were less astute in setting goals that matched the lacks in their planning behaviour.

By these many self-regulatory factors as well as their L2 achievement, the Achieved participants did indeed seem to fit the non-gainer category. Even so, some aspects of their behaviour seemed to reveal a good level of motivation. While two thirds of Achieved participants reported more difficulties with formulating than other participants, they also reported more strategies to overcome those difficulties; two-thirds set goals to improve their planning; and while only 40% set goals to improve their formulating, it was the highest percentage of any group. Half of them also reiterated their goals in the final open-ended section of their reflections. However, they were the least likely to set goals to continue learning French the following year (only two intended continuing).

It seems that these Achieved participants set performance goals (those related to demonstrating competence compared with others) rather than mastery goals (those related to developing competence). Oxford (2011) indicated that students who set mastery goals have stronger motivation, choose more cognitively challenging goals, and persist longer. Schunk and Zimmerman (2008) found that students who set mastery goals not only used more effective strategies, and more often, but they also "recovered more quickly from poor performance on the first exam in the course and displayed higher performance by the end of the course" (p. 7). It does seem, therefore, that, despite their motivation to succeed in the NCEA examination, the Achieved participants were not able to increase their learning based on written self-talk alone. They needed explicit L2 support and instruction in spending time relative to the task, in strategy use and in goal setting, as well as explicit language instruction, before being able to benefit from the reflective auto-input of the intervention.

To summarise, there appear to be strong links between the self-regulatory practices reported in the L1 reflective blogs and the L2 achievement levels of the participant groups. There also appear to be strong links between L2 achievement levels and continued learning. This leads to addressing the final research question. The L1 reflections offered several more insights that might inform teaching and learning; these observations are introduced in this chapter and lead to the recommendations listed in the final chapter (see Section 8.4).

The first three observations arose from an analysis of the participants' planning behaviour. Firstly, one third of participants regretted not spending enough time planning for the writing test. Because time spent approximated achievement for participants, allowing more time during the planning period is likely to have improved achievement levels. Secondly, 11% of participants listed memorising a script as a planning strategy, which was one of the concerns expressed when offering the participants seven days to prepare for the writing test with full access to L1 task instructions. Long-term mastery goals were clearly forgotten as participants focused on the test rather than the language they needed to learn for the test. A third observation was that, while 75% of participants self-observed and self-judged planning deficits accurately, their self-reactions were not so astute. Most Merit and Achieved participants needed help setting appropriate goals for overcoming the lacks in their prior planning behaviour.

The fourth observation related to the formulation process. All but two participants (97%) were able to list examples of the difficulties they experienced when formulating during the test and just over half (56%) were able to report the strategies they used to overcome these difficulties. However, only a third (35%) were able to report composition successes, and only a third (38%) were able to set goals for improved composing, and most of these related to planning, not composing. With inadequate planning, much of the important L2 formulating stage of composition would have occurred during the test rather than before it which would have caused all kinds of difficulties (Graham & Macaro, 2007).

The fifth observation related to the participants' goals for continuing to learn. Fifty-six percent of participants received the same grade for the first draft of the first assessment task in Term 1 as they received for their entire NCEA writing portfolio at the end of the academic year. While the grades awarded for the first drafts did not appear to influence decisions to continue learning, the final NCEA portfolio grade did. It seems that there was a high correlation between the extrinsic

motivation gained from the final NCEA grade, which indicated their overall level of achievement in L2 writing, and intentions to continue learning the following year. It was interesting to note, however, that while the final NCEA portfolio grade appeared to be instrumental in encouraging participants to continue learning, the indicator grades awarded on the first drafts of each task were not. Participants clearly understood they were able to improve those grades; their self-efficacy and motivation seemed high.

7.3: Three Perspectives

The primary interpretation of the research findings, as expounded above, is the interactive-sociocultural interpretation that framed the research. However, two further perspectives are of interest, a Skill-building perspective, and a metaphorical perspective. Skill-building Theory can be used to draw parallels between L2 acquisition and strategic self-regulated learning. Macaro (2001) noted that the knowledge of strategies, like L2 knowledge itself, moves from declarative to procedural through practice. Zimmerman (1990) referred to the procedural knowledge of strategies as routinised performance, which parallels fluent L2 use. Macaro, et al. (2016) referred to a dynamic continuum between linguistic knowledge and strategic knowledge. As linguistic knowledge becomes more fluent, less strategic knowledge is required, and as new declarative L2 knowledge is added, more strategic knowledge is required.

To extend the argument further, the reflection intervention in the current research equates to procedural knowledge. If practised over time, declarative reflection will eventually become procedural reflection and will occur simultaneously during language learning and without the need of dedicated time and the support of a template. Until this happens, reflecting and recording regularly, just like repeated L2 output, will ensure "a more durable memory trace" (Izumi, 2002, p. 570).

It is possible that, from a Skill-building perspective, the Achieved participants needed more support in the form of explicit explanations and tutoring as well as more practice in developing declarative reflective practices in relation to specific language features. The Excellence participants, on the other hand, had already developed procedural reflective practices, so they arguably had no need of, nor derived any benefit from, the dedicated time and template for some language features yet needed more reflective practice for others.

A third way to view the self-regulatory process of language learning is through the metaphor of a 'bi-cycle'. Figure 14 represents the reflective writing-to-learn-language journey. It rests against a background composed of the mediated learning experience (Feuerstein & Feuerstein, 1999) and its essential mediations of intentionality-reciprocity (dialogic engagement between teacher and learner), transcendence (which is strongly linked to goal setting and achieving and involves understanding the transferability of the present learning experience), and meaning (which is also a primary emphasis in communicative language teaching methodologies).

Figure 14: Bi-Cycle Metaphor for the Reflective Writing-to-Learn-Language Journey

Reflective Practices and the Feuersteinian Mediated Learning Experience:
On a Background of Intentionality-reciprocity, Transcendence and Meaning

The Teacher:
Provides direction and guidance with the tools and scaffolding of time and template to reflect

The Learner:
Chooses the direction and speed

Motivation and self-efficacy:
Propel the learner forward

ZPD-Interlanguage Development Road

ZPD-Interlanguage Development Road

The Self-Regulation Cycle:
- *self-observation*
- *self-judgment*
- *self-reaction*

as evidenced in the L2 writing and in the L1 reflective blogs and the processes, strategies and goal-setting behaviours reported

The Cognitive, Linguistic Cycle:
- *noticing-triggering*
- *hypothesis-testing*
- *metalinguistic reflection*

as evidenced in the L2 writing and in the L1 reflective blogs and the processes, strategies and goal setting behaviours reported

(Bicycle image from dbclipart.com)

The mediation in the Figure 14 image is that of the teacher providing the learner with the tools and scaffolding required to reflect in the first language while writing in the second. Based on many extrinsic and intrinsic factors, the learner chooses the direction and speed of the learning along the ZPD-Interlanguage Development Road (Williams & Burden, 1997) and is propelled forward by motivation and self-efficacy.

As the bicycle moves forward, the wheel of self-observation, self-judgment and self-reaction turns (Schunk & Usher, 2013) and causes the cognitive-linguistic wheel of noticing-triggering, hypothesis-testing and metalinguistic reflection to turn also (Swain, 2006). Evidence that these wheels are turning is found in the L2 writing and the L1 reflective blogs of the participants in this study. Because different language features develop at different rates for different learners (Benevento & Storch, 2011; Torras & Celaya, 2001), this journey is not a linear one but is dynamic and discursive, with many twists and turns over time.

It must be noted, also, that the Assessment *for* Learning writing portfolio structure is an environment that encourages these self-regulatory and linguistic cycles. As students plan to write, as they formulate ideas and transfer them to language in their first drafts, as they receive and respond to teacher feedback and participate in their own learning through auto-feedback or inner speech, as they choose remedial practice exercises, and as they rewrite multiple drafts over time, the self-regulatory behaviours of self-observation, self-judgment and self-reaction, and the language-learning behaviours of noticing-triggering, hypothesis-testing and metalinguistic reflection are more likely to develop and so the capacity to learn language is increased.

7.4: Summarising the Chapter

This chapter set out to synthesise the results of the L2 data analysis with the findings of the L1 data analysis in order to answer the research questions. To address the first research question (Does L1 written reflection increase L2 development when adolescents write for portfolio assessment? If so, what is the type and amount of increase and for whom?), the study found that the reflection intervention had a medium to large impact on the L2 written proficiency and on the accuracy of the Term 1 Merit participants. It had no significant impact on their complexity, fluency, verb error elimination or verb self-regulation. It also appeared to have no significant impact on the Term 2 Merit group or the Achieved and Excellence groups. It seemed as if the Term 1 Merit group, in particular, equated to Budoff's (1968) 'gainers' or Sasaki's (2000) 'novices before instruction' – they were at a point in their academic year and in their ZPD where they were ready to take next steps in interlanguage development related to L2 written proficiency and accuracy and the reflection intervention helped them to do so, but they needed more explicit support before being able to develop their verb knowledge. The Achieved participants equated to Budoff's 'non-gainers' or Sasaki's 'novices after instruction' and the Excellence participants to Budoff's 'high

scorers' or Sasaki's 'experts'. Further insights and confirmation of this interpretation were sought in the L1 blogs.

To address research question two (What insights, if any, does the content of the L1 reflections reveal concerning the results of the L2 data analysis?), findings appeared to confirm the three-level interpretation above. The Excellence participants indicated in their L1 blogs that they were more likely to spend more time planning for the writing test, to memorise more, to use more effective strategies, to use those strategies in combination, to be more aware of composition successes, to be more aware of their typical error types, and to need less help linking goals to previous behaviour. These students were at a stage in their L2 written proficiency and their accuracy where they required less input from the teacher and were more fluent in NZC Level 5 and 6 language use (see Section 1.2.2). Their L1 private speech was as fluent as their L2 written proficiency and accuracy. However, like other participants, they needed further support, in addition to reflection, in order to progress in verb knowledge. It is possible that the assessment writing portfolio process with its corrective error feedback inhibited growth in complexity and fluency.

The Term 1 Merit participants who gained most from the intervention were, by every measure of their L2 writing and L1 reflections, closer in performance to the Achieved participants than their Term 2 counterparts, who were closer to the Excellence participants. Their self-judgment and self-reaction as evidenced in their goal setting behaviour indicated their high motivation to improve. It seemed that these participants were at a stage in their ZPD where they benefitted from the dedicated reflection time and guidance and were able to take bigger steps in interlanguage development than would have been possible without these mediations. It seems that the reflection acted as a written form of private speech and, as such, activated the cognitive functions of noticing-triggering, hypothesis-testing and metalinguistic reflection which, in turn, increased their capacity to learn language.

The Achieved participants, however, were not yet ready to take advantage of reflective practices to improve their language learning because they still needed the additional mediation of explicit support in order to progress. What is gratifying, nonetheless, is that, while the Achieved participants expressed the most dissatisfaction with their planning behaviour, they set the most goals to improve their planning; and while they reported the most difficulties while composing,

they also reported the most strategies to overcome those difficulties. This runs counter to claims

that strategy use equates to achievement level. It also indicates the motivation of these

adolescents to achieve in the NCEA writing assessment.

In answer to the third and fourth questions (Are there any links between the reflective self-

regulatory practices evidenced in the blogs, achievement levels, and/or continued learning? and

What further insights regarding language learning do the blogs provide?), the study found strong

links between the self-regulatory practices reported in the L1 reflective blogs and the L2

achievement levels of the participant groups as well as strong links between L2 achievement

levels and continued learning.

Table 45: Further Insights with Teaching Practice Suggestions

Insight	Recommendation
1. One third of participants regretted not spending enough time planning for the writing test and time spent approximated achievement level.	1. Add 'making time relevant to task demand' to strategy lists.
2. While 75% of participants self-judged planning behaviour accurately, most Merit and Achieved participants needed help linking planning goals with prior behaviour.	2. Provide support to lower achieving students in linking learning goals to prior learning deficits.
3. Eleven percent of participants listed script memorisation when planning for the writing test.	3. Ensure students understand how to learn language when learning for writing tests. Provide fewer task details in advance of writing tests. Explore alternatives to tests in process writing units.
4. While 98% of participants listed formulation difficulties during the test, 56% reported strategies to overcome these difficulties, only 35% reported composition successes, and only 38% set goals for improved composing, yet most of these related to planning.	4. Develop training specific to the composition (linguistic formulation and physical production) process of L2 writing.
5. Fifty-six percent of participants received the same grade for the first draft of the first assessment task in Term 1 as they received for their entire NCEA writing portfolio at the end of the academic year. While the grades awarded the first drafts did not appear to influence decisions to continue learning, the final NCEA portfolio grade did.	5. Use other forms of measurement and assessment in addition to the NCEA rubric to guide teaching and learning throughout the year (see quadrant of measures in Section 4.3.5)

As indicated in Table 45, further insights provided by the blogs resulted in five recommendations:

1. that 'making time relevant to task demands' be added to strategy lists to emphasise the

importance of making time to teachers and learners; 2. that adolescent learners be offered more

guidance in linking learning goals with lacks in learning behaviour; 3. that teachers ensure their students understand the importance of learning language for acquisition as opposed to learning whole scripts for tests, and that, where possible, they provide fewer task details in advance of writing tests and explore alternatives to tests in process writing units; 4. that teachers develop training units that focus specifically on the composing (linguistic formulation and physical production) process of L2 writing; and 5. that forms of assessment, additional to the NCEA rubric, be used to guide teaching and learning throughout the academic year.

This chapter set out to synthesise the results of the L2 data analysis with the findings of the L1 data analysis in order to answer the research questions. The conclusions drawn naturally lead to the recommendations for teaching practice above. These are expanded further in the next and final chapter of this thesis.

CHAPTER EIGHT: LOOKING BACK AND MOVING FORWARD

The previous chapter synthesised and interpreted the L1 and L2 data analyses. This final chapter looks back at the pedagogical motivations behind the research and considers how these have been addressed by the study thus far. It also outlines several limitations to the study and considers implications for practice and directions for future research.

8.1: Pedagogical Motivations

The research questions posed in this study arose from three issues encountered in my own teaching practice, student retention, metacognition, and the measurement of learning. The first issue, retention, specifically relates to the retention of foreign language students into senior secondary school classes. As a FFL teacher in Aotearoa New Zealand, along with my colleagues, I have long been concerned about the diminishing numbers of students learning languages in senior classes. Together we have wondered what we might do to change the pattern. The second issue was higher order thinking, metacognition and self-regulation, what they look like in foreign language settings, and whether it is worthwhile teaching them when time is tight, especially in portfolio assessment contexts which are time-consuming but conducive to reflective processes. Despite being tempted to relegate reflective processes to out of school hours, I learned quickly from my students that, unless class time was allocated for them to reflect, too many did not make the effort to complete reflection exercises given for homework or even to improve their writing in response to the feedback they had been given. The third issue was the collection of fine-grained data capable of informing daily teaching and learning. An ERO (2012) report stated that two-thirds of New Zealand schools were not effective in differentiating practice based on formative assessment data. This was confirmed by Hattie and Yates (2013) who found that teachers have many tools to measure summative achievement but few tools to measure the developmental steps along the way that allow them to support daily teaching and learning in the manner that the ERO expected.

Motivated by the desire to use data to inform my practice and by the hunch that making time for my students to reflect on their language learning would lead not only to improved learning but also to improved motivation and retention, I conducted informal research with one cohort of my students over a two year period. As my students reflected on goal setting and achieving, and on the processes and strategies they used for internal and external assessments, I tracked their

achievement and retention rates and compared them with previous cohorts. I found superior achievement in national assessments at the ends of Year 11 and 12, and an unusually high retention rate from Year 11 into Year 12 and from Year 12 into Year 13, in fact, the best in my subject and in my school over the previous eight years. I therefore embarked on the present more formal research to understand more fully the impact of reflective processes on language learning in writing portfolio assessment contexts.

As I began my research, I discovered that studying the impact of reflective practices in foreign language contexts is not common. I was encouraged and motivated by Belgrad (2013) who called for empirical studies into "the belief that engagement of students in reflective learning processes has the potential to increase achievement, communication (voice), and motivation" (p. 12). I was also encouraged and motivated by Ortega (2009) who called for research that adds knowledge to the L2 writing field by focusing on "a wide range of school, university, workplace, and virtual settings across diverse geographical and institutional (foreign language) contexts" (p. 251). I was particularly encouraged and motivated by Manchón (2011a) who called for more research that worked at the interface of the SLA and L2 writing fields, an area she called writing-to-learn-language. My research was therefore designed not only to answer my own questions but also to respond to the above challenges. It responded to Belgrad (2013) by investigating whether reflective learning processes increase achievement in L2 learning contexts; it responded to Ortega (2009) by contributing to the L2 writing field with knowledge on L2 writing in the very specific New Zealand secondary school portfolio assessment context; and it responded to Manchón (2011a) by working at the writing-to-learn-foreign-language interface of the SLA and L2 writing fields.

8.2: Research Design

The research was a mixed method, counter-balanced, comparative study that asked whether first language written reflection increases second language learning when adolescents write for portfolio assessment. In a four-week process writing unit held at the end of Term 1, two French classes from two coeducational secondary schools took part in the reflection intervention while three French classes in two other schools in a different part of the city did not. Then in a second process writing unit at the end of Term 2, the Term 1 comparison (no reflection) group became the Term 2 intervention (reflection) group and vice versa. During the writing units, the participants

were given the task instructions one week in advance of a timed writing test. One week after the writing test they received feedback before writing a second draft for homework. Those taking part in the intervention were prompted by a template to record the strategies they used as they planned, composed and reviewed their work. Defining L2 learning as the changes made to second drafts as a result of feedback on first drafts, the difference between the changes made when participants reflected and when they did not was calculated. To facilitate a more fine-grained comparison of the gains with and without reflection, the intervention and comparison groups were further divided into NCEA subgroups, Achieved, Merit, and Excellence, using an adapted NCEA rubric. Any statistically significant gain after reflection was considered to possibly result from the reflection intervention.

8.3: Results, Findings and Interpretation

The goal of the intervention was to investigate whether those participants who were given the time and the occasion to reflect in English as they wrote in French had a greater gain in L2 learning (as evidenced by the changes in their second draft) than those who were not given this opportunity. Gains were measured using situated versions of East's (2008) Scoring Rubric, CAF analysis, error analysis, and Aljaafreh and Lantolf's (1994) Regulatory Scale. From the statistical analyses conducted, it appeared that the Merit intervention participants, in particular, the Term 1 Merit intervention participants, increased their L2 learning in a statistically significant manner, and that they did so in terms of L2 written proficiency and also in terms of accuracy (but not in terms of complexity, fluency, verb error elimination or self-regulated verb knowledge).

The L1 written reflections were also analysed to see whether they might offer explanations for these results. By comparing the L1 reflection data of the Term 1 Merit intervention group with that of the Term 2 Merit intervention group, then comparing the L2 writing data alongside the reflection data, it became apparent that, by every measure used in the research, the Term 1 group were at the lower end of the Merit band and the Term 2 group were at the higher end. Because researchers have found a strong link between self-regulatory strategy use and achievement levels, the Term 1 Merit participants clearly had more scope for microgenetic development. The strategies they reported using to overcome composition difficulties as well as their goal setting behaviour indicated their motivation to participate in the self-regulatory learning process and to succeed in the writing portfolio assessment.

This group were equivalent to Budoff's (1968) 'gainers' or Sasaki's (2000) 'novices before instruction'. The Excellence group did not require the mediation of the reflection intervention to progress in L2 written proficiency and accuracy; they were equivalent to Budoff's 'high scorers' or Sasaki's 'experts'. The Achieved group required much more than dedicated reflection time in order to progress. It seemed they were not ready to take advantage of the reflection intervention because they needed the mediation of explicit language support in addition to reflection in order to move forward. They equated to Budoff's 'non-gainers' or Sasaki's 'novices after instruction'. However, it was apparent that all participants required more specific support in addition to dedicated time and scaffolding for reflection before being able to improve in verb knowledge. It seems that the assessment context, in particular, the limited feedback permitted in NCEA writing portfolios, may have been instrumental in inhibiting the participants' growth in complexity and fluency.

8.4: Implications for Practice

Based on the analysis of the data, several recommendations for teaching practice are able to be offered. In echo of the words written by Ellis (2005), rather than prescribing what teachers should do, this research hopes to stimulate thought and inspire teachers to try new reflective approaches according to their local conditions.

The study asked about the impact of L1 reflection on L2 learning when adolescents write for portfolio assessment. In so doing, it drew attention to the links between reflective self-regulatory practices, writing to learn language, and Assessment *for* Learning. It framed the study within sociocultural theory, which allowed the students' reflective blogs to be viewed as auto-input and private or inner speech, and the time set aside for reflection and the template to guide reflection to be viewed as scaffolding which supports not only the learning of language but also the acquisition of reflective, self-regulatory practices.

The study found a strong correlation between the students' strategic, self-regulatory practices and their L2 achievement levels. Prior scholarship has found that, when strategic, self-regulatory practices are taught and applied, achievement levels rise (Cohen & Macaro, 2007; Feuerstein, et al., 1991; Graham & Macaro, 2007; Grenfell, 2007; Guo, 2012; Oxford, 2011; Zimmerman, 1990). As a consequence, teachers might consider setting aside time in their programmes to teach strategic, self-regulatory practices and to encourage self-recording.

Following the recommendations of Graham and Macaro (2007), foreign language teachers might begin their L2 writing programmes with a diagnostic reflection and writing exercise in order to establish which strategies individual students are using well and which strategies might help them increase their language learning and grades. Strategy awareness instruction tailored to students' needs might then follow. Teachers might also like to consider three issues that have arisen from this current study: firstly, the strategic use of time when planning to write; secondly, the formulation stage of writing that students find most difficult; and thirdly, the setting of improvement goals that relate specifically to prior writing deficits.

Firstly, drawing students' attention to time-related issues is important. For example, the study found that achievement levels approximated the amount of planning time spent. Adding 'making time relevant to task demands' to strategy lists might encourage students to spend appropriate amounts of time when planning. This study found that those who spent more time planning to write also spent more time learning (memorising/ remembering) language. This led to more fluent writing under test conditions which, in turn, led to more time for editing and improving writing and for the use of strategies that indicate critical sensitivity to language. Roca de Larios, et al. (2008) found that less proficient writers spend more time than more proficient writers converting ideas into language, and less time monitoring and evaluating their work. If students understand the importance of time spent when planning to write, they might experience improvements in writing proficiency as well as language learning outcomes. They could also be reminded, however, of the transcendent or mastery goals of fluency as well as of the performance goal of examination success. This might inspire them to prepare by learning language rather than rote learning entire scripts.

Secondly, helping students to understand the complexities of composing in a second language (formulating) could be added to strategy instruction. Formulating involves converting ideas to language and physically producing a draft with purpose and audience in mind and is a more effortful exercise than composing in a first language. The current study found that, while all but one student reported difficulties when composing, only about half reported strategies to overcome difficulties. Based on their own research which found that few adolescents plan when writing, Cohen and Macaro (2007) suggested teaching students to brainstorm or mind map in the target language so as to retrieve language from long-term memory, then to evaluate the gap between what they wanted to say and what they were able to say in their first language. This might involve

combining one set phrase with another, restructuring a set phrase by changing parts of it, translating, consulting a bilingual dictionary, or even avoiding the sentence altogether. Students might also be reminded to monitor or check their writing using strategies such as visual monitoring, auditory monitoring, backtranslating, or prompting with a reminder, a rhyme or a mnemonic.

Thirdly, teachers might consider including in their strategy instruction a component which helps students set improvement goals that relate specifically to their prior writing deficits. This study found that, although two-thirds were already aware of their typical error types when writing, only one third reported goals to improve future writing, and most of these related to planning rather than formulating. It also found that only a third articulated composition successes. In order to improve written proficiency and language learning, students must be encouraged to enunciate what has gone well, so they might repeat their successes, in the same way that they are encouraged to note what is lacking so they might make improvements. In order to become more self-regulated they also need to be offered choices. Offering students a choice of remedial exercises based on their most common lacks in proficiency provides scaffolding for future writing events when they might need to find such exercises themselves. Involving students in some form of error analysis is a good way to do this. Because errors are an integral feature of interlanguage development, error analysis has the potential to encourage the learner to test out hypotheses, respond to feedback, to choose remedial practice exercises related to their most frequent errors, and to discover language rules in the process (Ellis & Barkhuizen, 2005). In this study, two-thirds of students found the error coding feedback helpful in improving their language and their writing whereas just over a third found the other types of feedback to be helpful.

In process writing contexts, feedback is an important component. In addition to the strategy training suggested above, teachers might consider the differences between feedback for acquisition and feedback for accuracy. Manchón, (2011a) described feedback for acquisition as feedback which prompts deep linguistic processing and which is most likely to occur in writing programmes where learners receive differentiated feedback specific to their ZPD, are given time to reflect on this feedback and to notice the gap between their own knowledge and this new knowledge, and to produce this new knowledge in subsequent drafts. Kuiken and Vedder (2008) concluded from their research that the depth of the metalinguistic processing that occurs while students write, whether simple or elaborate, will determine the significance of their L2 outcomes.

In portfolio assessment contexts where explicit teacher feedback is limited by examination exigencies, adolescents would benefit from developing the skills of self-observation, self-judgment and self-reaction, so they understand, based on their previous writing, the steps they need to take to push their writing proficiency and language learning forward.

Finally, teachers might consider developing a toolbox of more fine-grained forms of measurement (in addition to the rubrics typically used in portfolio writing contexts) that are capable of exposing the small, dynamic, daily changes in language that indicate learning is taking place. Learners who are given opportunities to participate in the development and use of such measures alongside opportunities to reflect on their usefulness for learning are more likely to understand their personal learning needs and to develop their ability to self-observe, self-judge and self-react. Because different language skills develop at different rates for different students, the use of these alternative measures has the potential to provide teachers as well as learners with the kinds of data that are able to inform the teaching and learning process throughout the year (see the quadrant of measures in Section 4.3.5) as well as motivate learners to continue learning as they recognise and participate in their progress over time.

8.5: Limitations

There are several limitations to the usefulness and transferability of the findings of this study. The first limitation relates to the very specific research context, the Year 11 FFL classrooms in the New Zealand secondary schools where the participants were writing for NCEA portfolio assessment. While the specificity of the context might limit generalisation to some contexts, it is also an asset (Benevento & Storch, 2011; Manchón, 2011a) as it provides data from real students in real classrooms preparing written samples for a real, high-stakes portfolio writing assessment, a type of examination used often in communicative contexts around the world, for example, in Australia (Benevento & Storch, 2011), in Britain (Graham & Macaro, 2007), and in Hong Kong (I. Lee, 2011). Ortega (2009) argued that there is a need for research to be conducted in various specific contexts around the globe, such as this study, so that knowledge in the L2 writing field might be increased and theory and policy developed.

A second possible limitation related to context is the fact that the research took place in classrooms where the data were able to be used for national assessment purposes. This made the random sampling that is desirable in more experimental research difficult, if not impossible.

Due to the nature of secondary school timetables and the optional status of language learning, the selection of the intervention and comparison groups needed to be by convenience rather than randomisation; whole classes became either intervention or comparison groups, and the schools in one geographical region became the intervention group while the schools in another geographical region became the comparison group. Internal validity may have been weakened as a result of not using randomisation (Nunan, 1992), but the data can still be called clinically-elicited and have psycholinguistic validity because they consisted of authentic language which was produced in a communicative context where meaning and message were the primary concerns and which was therefore typical of real language (Ellis & Barkhuizen, 2005).

A third context-related limitation is that the data collection process did not always go as planned, which is often the case when researching real students in real classrooms for real purposes (Ellis & Barkhuizen, 2005). Of the 95 students who were initially available to take part in the research, only 71 completed all steps in the ethics approval process, wrote all four drafts, and completed the two-stage guided reflection that formed the intervention. In addition, there were interruptions. Some students in the class where the first intervention was conducted struggled to complete their reflection during class time because they were late to class after lunch, had issues with the slow school network, or had a malfunctioning computer which necessitated starting over. They were therefore asked to complete the reflection for homework rather than in class. The lessons learned regarding the time taken for students to arrive after lunch, to log into networked computers, and to complete online reflections ensured that interventions undertaken in subsequent classrooms were conducted in a more efficient and consistent manner. In addition to these issues, in every lesson there were one or two absentees, a usual pattern of school attendance. These absent participants were also asked to complete their compositions or reflections on another occasion but under similar conditions. Given the above, only 71 of the 95 participants finally completed the writing of the L2 drafts and the L1 blogs as planned. This meant that the intervention and comparison groups were smaller and more unbalanced in number than originally planned. The data from the small sample size was strengthened in reliability by 'triangulating' the data collection and the data measurement processes; to this end a quadrant of measures was developed (see Section 4.3.5) so both the L1 and L2 data was collected and measured using a mix of objective tools and tools that require expert judgment. Any imbalance was also taken into consideration when reporting statistical significance computations but meant that, in some cases, the assumption of

homogeneity of variance was violated. In these cases dot plots were used to make the findings more transparent for the reader.

A fourth context-related limitation resulted from the need to conduct fair assessment and not advantage or disadvantage any participant over another. This ethical consideration necessitated a counterbalanced research design but, in so doing, the possibility of a 'carryover' effect was introduced. Because the Term 1 intervention group became the Term 2 comparison group, the Term 2 comparison group could very possibly have retained advantages in reflective learning as a result of having already taken part in the reflection intervention in Term 1. It needs also to be noted, that, while the high-stakes NCEA assessment context might have forced the research design to some degree, it also ensured the fifteen-year-old participants' sincere and genuine engagement with the research process.

A fifth limitation relates to the fact that the two drafts of each writing unit were not written under the same conditions. The first draft was written as a test without access to resources and the second draft was written as homework and with access to resources. However, participants had access to the L1 task instructions seven days in advance of the test and were therefore able to evaluate and overcome the gaps in their language knowledge with access to resources prior to the timed test. In addition, both intervention and comparison groups received the same treatment. It seems that the two writing conditions did not appear to affect the type of language changes that occurred in the second draft. Because NCEA students and teachers must guarantee that the writing presented for assessment is the student's own work, first drafts are often written under supervision and in class. In addition, the NCEA writing standard states that subsequent drafts may be improved over time and with access to resources. However, students are aware that any changes they make to their second drafts must bear resemblance to their first draft if they wish to maintain proof of authorship. They are therefore less inclined to add content. They are further discouraged from adding content by the fact that they can receive feedback only once (NZQA, 2016). Anything added to the second draft will therefore not receive feedback. The question remains. Can changes made when participants have access to resources be considered L2 learning? I would argue with others that the cognitive processes activated by rewriting, whether with or without access to resources, are indeed part of the L2 learning process (Bitchener & Storch, 2016; Ferris, 2006; Gass, 1997; Manchón, 2011b; Polio, 2012). As participants located the language forms indicated by the indirect coded feedback, then noticed the gap between what

they had written and L1 writer norms, perhaps researching and practising the form in order to do so, then as they used the edited form in their rewriting, they were actively taking part in the L2 learning process (Ellis & Shintani, 2014; Swain & Lapkin, 1995.)

While the five previous issues resulted from the classroom and assessment context in which the research took place, all had positive as well as negative aspects. The sixth limitation relates to the 'dosage' or amount of reflection that was offered the participants and could lead to further research. Reflection is a learning habit that can be developed over time and a blog is an online journal for recording regular thoughts over time. The current research conducted a single event reflection intervention and blog entry, albeit in two phases. This single event provided a statistically significant impetus to the L2 learning of the Term 1 Merit participants only. If the reflection had occurred, as would have been ideal, regularly and over time, it is possible that the Achieved and Excellence participants might have gained as well as the Merit participants and the gains might have occurred in verb knowledge, complexity and fluency as well as in L2 written proficiency and accuracy. This assertion is made on the basis of Muijs (2011) who advised that research design requires careful consideration regarding the right level of treatment. In this respect it is similar to a doctor prescribing the right level of medicine. The right dosage can be effective; too little will have no effect; and too much could be fatal, or as Muijs put it, "[w]hile the consequences of too much educational intervention are usually less serious, getting the 'dosage' right is nonetheless important" (p. 17). In this research it is possible that the dosage was too small and did not continue for long enough so the results were somewhat diluted.

8.6: Moving Forward

Despite the difficulties and constraints posed by conducting research in real classrooms with real students for real purposes, research must continue to take place in foreign language classrooms and into foreign language writing for assessment purposes. This context differs from writing in first language classrooms because students are writing to learn language rather than learning how to write. Research in this area contributes to developing a "situated understanding of writing practices in (foreign language) instructional settings" (Manchón, 2009, p. 4) and adds new knowledge to a relatively new field of research.

To add to the insights gained as a result of this current research, and in response to Muijs (2011) 'dosage' argument, further research might be conducted into the processes, strategies and goal

setting behaviour of secondary school students over a three-term period, (that is, virtually the whole academic year), starting early in Term 1 and continuing through to the end of Term 3. This would allow for a descriptive study, or, indeed, another intervention study (if participants willing to take part were able to be located), in which students record entries in reflective blogs more regularly than in the current research, and related to every script written for assessment or for practice during the time period. This would allow for a more appropriate 'dosage' of reflection and the development of reflective habits which would perhaps reveal the true (or additional) benefits of reflective learning. It is possible that a more longitudinal study might result in students from all three achievement groups gaining in a statistically significant manner and in more language features than the current research revealed.

In addition, research could investigate the impact of the recommendations to teachers in Section 8.4 above: the findings of a diagnostic reflection exercise to discover which strategies students are using well and which might help them increase their language learning and grades; the impact of strategy training specifically related to process writing - planning time, formulation, and goal setting; the impact of feedback designed for acquisition rather than accuracy; and the impact on language learning of the development and use of a toolbox of formative measuring tools.

Research might also investigate whether the sort of indirect feedback on errors allowed in portfolio writing for national assessment is more effective for language learning than explicit feedback on error. Two thirds of the students in the current study expressed a view that it helped their language learning and subsequent their ability to improve their writing more than the holistic feed forward comment and indicator grade did, and more than the explicit teaching on common errors did. It is possible that the difficult cognitive wrangling that is required when receiving and responding to indirect coded feedback on error creates deeper memory traces than when receiving and responding to feedback that is easier because it is more explicit.

A different kind of future study might research New Zealand teachers regarding their process writing unit designs in order to ascertain how they apply Assessment *for* Learning principles when they are constrained by the NCEA writing assessment portfolio context, and how they ensure the authorship of their students' work in conditions other than timed tests.

8.7: Concluding Remarks

As a result of the investigations conducted here, I believe that, despite busy programmes, it is worthwhile for teachers to make time early in the school year for their students to begin developing reflective practices as they write for portfolio assessment. By regularly dedicating time and guidance to reflection, I suspect that language learning will be increased for all students, albeit at different rates, and with different language features.

In addition, I believe that this study underscores the strong links, not only between self-regulated learning and L2 written performance, as also confirmed through research by Teng and Zhang (2016), but between self-regulated learning, L2 written performance and Assessment *for* Learning. This is because L2 portfolio writing for assessment purposes provides unique opportunities for students to learn language as they produce multiple drafts over time, as they record and share their reflections on the processes and strategies they use while writing, as they receive and respond to indirect feedback on their writing, as they observe and judge the language they use as they write, and as they respond by setting goals to improve. Finally, when students are offered an opportunity to reflect in their first language as they write in their second language, they are able to communicate with their teachers regarding their learning in ways that they might not otherwise do, and thereby, they have the opportunity to build the kind of reciprocal learning relationships with their teachers that have the potential to increase their language learning over time.

REFERENCES

Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, I. (2009). Directions for assessment in New Zealand: Developing students' assessment capabilities. (Position paper). Wellington NZ: MOE.

Adams, R. (2003). L2 output, reformulation, and noticing: Implications for IL development. Language Teaching Research, 7(3), 347-76.

Alderson, J., Clapham, C., Steel, D. (1997). Metalinguistic knowledge, language aptitude and language proficiency. Language Teaching Research, 1, 93–121.

Aljaafreh, A., & Lantolf, J. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. Modern Language Journal, 78(4), 465-483. doi:10.2307/328585.

Anderson, J. (1993). Rules of the mind. Hillsdale NJ: Lawrence Erlbaum Associates.

Anderson, N. (2002, April). The role of metacognition in second language teaching and learning. CAL Digests, 1(10).

Anderson, N. (2008). Metacognition and good language learners. In C. Griffiths (Ed.), Lessons from Good Language Learners (pp. 99–109). Cambridge: Cambridge University Press.

Anderson, J., & Corbett, A. (1993). Teaching of Cognitive Skill. In Rules of the mind (pp. 235-255). Hillsdale NJ: Lawrence Erlbaum Associates.

Archibald, A., & Jeffery, G. (2000). Editorial: Second language acquisition and writing: A multidisciplinary perspective. Learning and Instruction, 10(1), 1-11.

Arievitch, I., & Haenen, J. (2005). Connecting Sociocultural Theory and Educational Practice: Galperin's Approach. Educational Psychologist, 40(3), 155-165

Assessment Reform Group. (2002). Assessment for learning: Beyond the Black Box. University of Cambridge School of Education (PDF). Available: http://www.nuffieldfoundation.org/assessment-reform-group

Bandura, A. (1977). Social learning theory. Englewood Cliffs, NJ: Prentice Hall.

Barnard, R., & Campbell, L. (2005). Sociocultural theory and the teaching of process writing: The scaffolding of learning in a university context. The TESOLANZ Journal, 13, 76-88.

Belcher, D. (2012). Considering what we know and need to know about second language writing. Applied Linguistics Review, 3(1), 131-150. doi:10.1515/applirev-2012-0006

Belgrad, S. (2013). Portfolios and e-portfolios: Student reflection, self-assessment, and goal setting in the learning process. In J. H. McMillan (Ed.), SAGE handbook of research on classroom assessment (pp. 330-347). (Online version). doi:10.4135/9781452218649.n19

Benevento, C., & Storch, N. (2011). Investigating writing development in secondary school learners of French. Assessing Writing, 16(2011), 97-110. doi:10.1016/j.asw.2011.02.001

Biggs, J., & Collis, K. (1982). Evaluating the quality of learning: The SOLO taxonomy. NY: Academic Press.

Bitchener, J. (2008). Evidence in support of written corrective feedback. Journal of Second Language Writing, 17, 102-118. doi:10.1016/j.jslw.2007.11.004

Bitchener, J. & Storch, N. (2016). Written corrective feedback for L2 development. Bristol: Multilingual Matters.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education, 5(1), 7-73. doi:10.1080/0969595980050102

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. Language Learning, 33(1), 1-17. doi:10.1111/j.1467-1770.1983.tb00983.x

Bruner, J. (1971). "The process of education" revisited. The Phi Delta Kappan, 53(1), 18-21.

Budoff, M. (1968). Learning potential as a supplementary testing procedure. In J. Hellmuth (Ed.), Learning disorders, 3. Seattle WA: Special Child.

Chamot, A., Anstrom, C., Bartoshesky, A., Belanger, A., Delett, J., Karwan, V., & Keatley, C. (2005). The elementary immersion learning strategies resource guide. Available: http://www.nclrc.org/eils

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. Journal of Second Language Writing 12(3), 267-296.

Chesterfield, R., & Chesterfield, K. (1985). Natural order in children's use of second language learning strategies. Applied Linguistics 6(1), 45-59. doi:10.1093/applin/6.1.45

Cimasko, T., & Reichelt, M. (2011). Foreign language writing instruction: Principles and practices. Anderson, SC: Parlor Press, LLC.

Claxton, G. (2002). Building learning power: Helping young people become better learners. Bristol: TLO.

Claxton, G. (2006, Sept. 6). Expanding the Capacity to Learn: A new end for education? Opening Keynote Address British Educational Research Association Annual Conference, Warwick University.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, A. (2011). Strategies in learning and using a second language. Harlow, England; New York : Pearson/Longman.

Cohen, A., & Macaro, E. (2007). Language learner strategies: Thirty years of research and practice. Oxford: Oxford University Press.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (electronic resource) (6th ed.). New York: Routledge.

Colby-Kelly, C., & Turner, C. (2007). AfL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes, 64(1), 9-37.

Corder, S. (1967). The significance of learner errors. International Review of Applied Linguistics (5), 161–170.

Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment (Online). Available: http://www.coe.int/t/dg4/linguistic/cadre1_en.asp

Cumming, A. (2001). Learning to write in a second language: Two decades of research. International Journal of English Studies, 2, 1-23.

Cumming, A. (2016). Theoretical orientations to L2 writing. In R. Manchón, & P. Matsuda (Eds.). *Handbook of Second and Foreign Language Writing, (*pp. 65-88). Boston: De Gruyter.

De Guerrero, M., & Villamil, O. (2000). Activating the ZPD: Mutual scaffolding in L2 peer revision. The Modern Language Journal, 84(1), 51-68.

De Keyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in second language acquisition 17*(3), 379–410. doi:10.1017/S027226310001425X

De Silva, R., & Graham, S. (2015). The effects of strategy instruction on writing strategy use for students of different proficiency levels. *System, 53,* 47-59.

Dörnyei, Z. (2001). *Motivational strategies in the language classroom.* Cambridge: Cambridge University Press.

Dyson, B. (2010). Learner language analytic methods and pedagogical implications. *Australian Review of Applied Linguistics, 33*(3), 30.1–30.21. doi:10.2104/aral1030

East, M. (2008). *Dictionary use in foreign language writing exams: Impact and implications.* Philadelphia, PA: John Benjamins Pub. Co.

East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing 14,* 88-115. doi:10.1016/j.asw.2009.04.001

East, M. (2016). *Assessing foreign language students' spoken proficiency: Stakeholder perspectives on assessment innovation.* Educational Linguistics series. Singapore: Springer

East, M. & Scott, A. (2011). Assessing the foreign language proficiency of high school students in New Zealand: From the traditional to the innovative. *Language Assessment Quarterly, 8*(2), 179-189. doi: 10.1080/15434303.2010.538779

East, M., Shackleford, N., & Spence, G. (2007). Promoting a multilingual future for Aotearoa New Zealand: Initiatives for change from 1989 to 2003. *Journal of Asian Pacific Communication, 17*(1), 11-28.

Education Review Office (2012). *Literacy and Mathematics in Years 9 and 10: Using Achievement Information to Promote Success.* Wellington: Education Review Office. Available: http://www.ero.govt.nz/publications/literacy-and-mathematics-in-years-9-and-10-using-achievement-information-to-promote-success/

Ellis, R. (1994). *The study of second language acquisition.* Oxford: Oxford University Press.

Ellis, R. (2005). *Instructed Second Language Acquisition: A literature review.* (Effective Pedagogy No. 5163). Auckland, NZ: Uniservices Ltd. Available: http://www.educationcounts.govt.nz/publications/schooling/5163

Ellis, R. (2008). Principles of Instructed Second Language Acquisition. *CalDigest.* Washington DC: Centre for Applied Linguistics.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language.* Oxford, NY: Oxford University Press.

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research.* New York: Routledge.

Evans, N., Hartshorn, K., Cox, T., & Martin de Jel, T. (2014). Measuring written linguistic accuracy with weighted clause ratios: A question of validity. *Journal of Second Language Writing, 24,* 33-50. doi:10.1016/j.jslw.2014.02.005

Excel (Version 14.0.7173.5000, 2010). Microsoft Home and Student (Software).

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in Second Language Writing: Contexts and Issues.* NY: Cambridge University Press.

Ferris, D. (2012). Written corrective feedback in second language acquisition and writing studies. *Language Teaching, 45,* 446–459. doi:10.1017/S0261444812000250

Feuerstein, R., & Feuerstein, S. (1999). Mediated learning experience: A theoretical review. In R. Feuerstein, P. Klein, & A. Tannenbaum (Eds.), *Mediated learning experience (MLE): Theoretical, psychosocial and learning implications* (pp. 3-48). London: Freund Publishing House Ltd.

Feuerstein, K., Klein, P., & Tannenbaum, A. (1991). *Mediated learning experience: Theoretical, psychological, and learning implications.* London: Freund Publishing House Ltd.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-11.

Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*, 365–387.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375.

Fotos, S., & Hinkel, E. (2007). Form-focused instruction and output for second language writing gains. In S. Fotos, & H. Nassaji (Eds.), *Form-focused instruction and teacher education: Studies in honour of Rod Ellis* (pp. 131-143). Oxford, NY: Oxford University Press.

Gal'perin, P. (1992). Stage-by-stage formation as a method of psychological investigation. *Journal of Russian and East European Psychology, 30*(4), 60-80.

Gardner, J. (1963). *Self-renewal. New York: Harper & Row.*

Garner, R. (2016, Aug 26). *The near-terminal decline of language-learning.* TES (Online newspaper article). Available: https://www.tes.com/news/school-news/breaking-views/why-its-finally-time-accept-we-need-just-one-exam-board

Gass, S. (1997). *Input, interaction, and the second language learner.* NJ: Lawrence Erlbaum Associates.

Gordon, L. (2008). Writing and good language learners. In C. Griffiths (Ed.), *Lessons from Good Language Learners (*pp. 244–254). Cambridge: Cambridge University Press.

Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Rating. (*Report for the US Department of Education). Center for Education Compensation Reform. Available: http://cecr.ed.gov

Graham, S. (2004). Giving up on modern foreign languages? Students' perceptions of learning French. *The Modern Language Journal, 88*(ii), 171-191.

Graham, S., & Macaro, E. (2007). Designing year 12 strategy training in listening and writing: From theory to practice. *Language Learning Journal, 35*(2), 153-173. doi:10.1080/09571730701599203

Grande, T. (2015). *Preparing data in Excel to import into SPSS.* (Video). Available: https://www.youtube.com/watch?v=NlCki6kqsdY

Grenfell, M. (2007). Language learner strategy research and modern foreign language teaching and learning, *The Language Learning Journal, 35*(1), 9-22. doi:10.1080/09571730701315576

Griffiths, C. (2008). Strategies and good language learners. In C. Griffiths (Ed.), *Lessons from Good Language Learners (*pp. 83–98). Cambridge: Cambridge University Press.

Grift, G. (2007). In J. Satchwell (Ed.), *Assessing the whole child: Creating powerful portfolios and student led conferences.* Heatherton, Vic.: Hawker Brownlow Education.

Gu, Y. (2010). (Review). In Oxford. R. (2011). *Teaching and researching language learner strategies.* Great Britain: Pearson Education Limited.

Guo, H. (2012). An empirical study on the relation between meta-cognitive strategies and listening autonomous learning ability. T*heory and Practice in Language Studies 2*(11), 2446-2451.

Hakuta, K. (2016) About Wordsift. (Web page). Available: https://wordsift.org/about.html

Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 241-266). Norwood, N.J.: Ablex Pub. Corp.

Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing, 8,* 5-16.

Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing, 11,* 329–350.

Harklau, L., & Williams, G. (2010). Practising theory in qualitative research on second language writing. In T. Silva, & P. Matsuda (Eds.), *Practising theory in second language writing* (pp. 93-111). West Lafayette, Indiana: Parlor Press.

Harley, B. (1989). Functional grammar in French immersion: a classroom experiment. *Applied Linguistics, 10*(3), 331–59.

Harley, B. (2007). The role of focus-on-form tasks in promoting child L2 acquisition. In C. Doughty, & J. Williams Eds.). *Focus on form in classroom second language acquisition,* (pp. 156-173). UK: Cambridge University Press.

Hattie, J., & Yates, G. (2013, August). Understanding learning: Lessons for learning, teaching and research (presentation). ACER Research Conference, Melbourne, Australia. Available: http://research.acer.edu.au/research_conference/RC2013/6august/10/

Houston, T., & Turner, P. (2007, Spring). Mindfulness and communicative language teaching. *Academic Exchange,* 138-142.

Hyland, K. (2009). *Teaching and Researching Writing.* (2nd ed.). Harlow: Longman.

Hyland, K. (2011). Learning to write: Issues in theory, research, and pedagogy. In R. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 17-35). Amsterdam, NLD.: John Benjamins Pub. Co.

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4*(1), 51–69. doi: 10.1016/1060-3743(95)90023-3

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativilisation. *Studies in Second Language Acquisition, 24*(4), 541-77. doi:10.1017/S0272263102004023

Johnson, D. (1993). Classroom-oriented research in second language learning. In A. Omaggio Hadley (Ed.), *Research in language learning: Principles, processes, and prospects* (pp. 1-23). Lincolnwood, Illinois: The National Textbook Company.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2,* 130-144. doi: 10.1016/j.edurev.2007.05.002

Kachru, B. (1992). World Englishes: approaches, issues and resources. *Language Teaching 25*(1), 1-14.

Knouzi, I., Swain, M., Lapkin, S., & Brooks, L. (2010). Self-scaffolding mediated by languaging: microgenetic analysis of high and low performers. *International Journal of Applied Linguistics, 20(1*), 23-49.

Krashen, S. (1995). *Principles and practice in second language acquisition.* Hertfordshire, England: Prentice Hall.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing, 17*(1), 48-60. doi:10.1016/j.jslw.2007.08.003

Lai, M. (2013, Sept10). *Quasi-experimental designs for demonstrating intervention success.* Quant-Dare Seminar Series, Faculty of Education, University of Auckland, NZ.

Lamy, M., & Goodfellow, R. (1999). Reflexive conversations in the virtual language classroom. *Language Learning and Technology,* 2(2), 43-61.

Lamy, M., & Hassan, X. (2003). What influences reflective interaction in distance peer learning? Evidence from four long-term online learners of French. *Open Learning, 18*(1), 39-59. doi:10.1080/0268051032000054112

Langer, E. (2000). Mindful learning. *Current Direction in Psychological Science, 9*(6), 220-223.

Laoire, M. (2014). Learners' reflections on their narratives on L2 and L3 learning. In *Essential topics in applied linguistics and multilingualism.* Second Language Learning and Teaching (pp. 235-251). Switzerland: Springer. doi: 10.1007/978-3-319-01414-2_14

Lantolf, J., & Beckett, T. (2009). Sociocultural Theory and second language acquisition. *Language Teaching, 42,* 459-475. doi:10.1017/S0261444809990048.

Lantolf J., & Poehner, M. (2014). *Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide.* NY: Routledge.

Lee, I. (2007). Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. T*he Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes, 64*(1), 199-214. doi:10.1353/cml.2008.0018

Lee, I. (2011). Issues and challenges in teaching and learning EFL writing: The case of Hong Kong. In T. Cimasko & M. Reichelt (Eds.), *Foreign language writing instruction: Principles and practices* (pp.118-137). Anderson, SC: Parlor Press.

Lee, I. & Coniam, D. (2013). Introducing assessment for learning in EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing, 22,* 34-50.

Lee, M.-K. (2015). Peer feedback in second language writing: Investigating junior secondary school students' perspectives on inter-feedback and intra-feedback. *System.* 55, 1-10.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*(2), 126-141. http://dx.doi.org/10.1093/applin/4.2.126

Long, M. (1991). Focus on form: a design feature in language teaching methodology. In *Foreign language research in cross-cultural perspective* (pp. 39–52)*. Amsterdam: John Benjamins.

Loveridge, J. (2010). Involving children and young people in research in educational settings: (Research and Evaluation by Victoria University of Wellington, Jessie Hetherington Centre for Educational Research; No. 80440). Wellington N.Z.: Ministry of Education. Available: http://www.educationcounts.govt.nz/publications/schooling/80440/downloads

Macaro, E. (2001). *Learning strategies in foreign and second language classrooms.* London: Continuum.

Macaro, E., Graham, S., & Woore, R. (2016). *Improving foreign language teaching: Towards a research-based curriculum and pedagogy.* UK: Routledge.

Manchón, R. (2009). Broadening the perspective of L2 writing scholarship: The contribution of research on foreign language writing. In R. Manchón (Ed.). *Writing in Foreign Language Contexts: Learning, teaching and research.*

Manchón, R. (2011a). Situating the learning-to-write and writing-to-learn dimensions of L2 writing. In R. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 3-13). Amsterdam, NLD: John Benjamins Pub. Co.

Manchón, R. (2011b). The language learning potential of writing in foreign language contexts: Lessons from research. In T. Cimasko & M. Reichelt (Eds.), *Foreign language writing instruction: Principles and practices* (pp. 44–64). Anderson, SC: Parlor Press.

Manchón, R. (2016). Introduction: Past and future of L2 writing research. In R. Manchón, & P. Matsuda (Eds.). *Handbook of Second and Foreign Language Writing, (*pp. 65-88). Boston: De Gruyter.

Manchón, R., Roca de Larios, J., & Murphy, L. (2009). The temporal dimension and problem-solving nature of foreign language composing processes, implications for theory. In R. Manchón (Ed.). *Writing in Foreign Language Contexts: Learning, teaching and research.*

Martella, C., Nelson, R., Morgan, L., & Marchand-Martella, N. (2013). *Understanding and interpreting educational research.* NY: The Guilford Press.

Martinez, C. (2012). Developing metacognition at a distance: Sharing students' learning strategies on a reflective blog. *Computer Assisted Language Learning, 25*(2), 199-212. doi:10.1080/09588221.2011.636056

Minister of Education. (1988). Tomorrow's schools: The reform of education administration in New Zealand. Wellington: Government Printer.

Ministry of Education (2007). *The New Zealand Curriculum.* Wellington: Learning Media. Available: http://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum

Ministry of Education (2007). Achievement objectives by learning area: Set of 8 charts. Wellington: Learning Media. Available: http://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum

Ministry of Education (2015). *Secondary subjects 1996-2015.* (Data file). Available: https://www.educationcounts.govt.nz/statistics/schooling/student-numbers/subject-enrolment

Ministry of Education (2016, 29 Nov.). *School Deciles.* (Web page). Available at: http://www.education.govt.nz/school/running-a-school/resourcing/operational-funding/school-decile-ratings

Moeller, A., Theiler, J., & Wu, C. (2011). Goal setting and student achievement: A longitudinal study. *The Modern Language Journal, 96*(2), 153-169. doi:10.1111/j.1540-4781.2011.01231.x

Moir Scott, A., & Tolosa, C. (2015). Locating the learning: A comparison of text-analysis tools in the foreign language writing of adolescents. *The New Zealand Language Teacher, 41,* 48-63.

Muijs, D. (2011). *Doing quantitative research in education with SPSS.* (2nd edition). London: Sage Publications Ltd.

Murray, L., Hourigan, T., & Jeanneau, C. (2007). Blog writing integration for academic language learning purposes: Towards an assessment framework. *IBÉRICA, 14,* 9-32.

Naiman, N., Fröhlich, M., Stern, H., & Todesco, A. (1978). *The good language learner.* Research in Education, 7. Toronto: Ontario Institute for Studies in Education.

New Zealand Qualifications Authority (2006). *Assessment Guidelines for NCEA Learning Languages 2006.* Available: http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/subjects/french/clarifications/writing-expiring-standards/

New Zealand Qualifications Authority (2015). *Standard 90882.* Available: http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2015/as90882.pdf

New Zealand Qualifications Authority (2016). *Languages clarifications.* Available: http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/subjects/languages/clarifications/1/writing/

New Zealand Qualifications Authority (2016, February 3). Reduction in the number of pieces of student evidence required for required Learning Languages portfolios. In *Assessment*

Matters (A2016/001). Available: http://www.nzqa.govt.nz/about-us/publications/newsletters-and-circulars/assessment-matters/reduction-in-the-number-of-pieces-of-student-evidence-required-for-required-learning-languages-portfolios/

New Zealand Qualifications Authority (2017, January 20). *The online curriculum guide.* Available: http://seniorsecondary.tki.org.nz/Learning-languages/What-s-new-or-different

Norris, J., & Manchón, R. (2012). Investigating L2 writing development from multiple perspectives: Issues in theory and research. In *L2 writing development: Multiple perspectives.* Berlin: De Gruyter Mouton.

Norris, J., & Ortega, L. (2005). Defining and measuring Second Language Acquisition. In C. Doughty, & M. Long (Eds.), *The handbook of Second Language Acquisition* (electronic ed., pp. 716-761). Malden, MA: Blackwell Pub. doi:10.1111/b.9781405132817.2005.00022.x

Nunan, D. (1992). *Research methods in language learning.* New York, NY: Cambridge University Press.

Nunan, D., & Bailey, K. (2009). *Exploring second language classroom research: A comprehensive guide.* Boston, MA: Heinle, Cengage Learning.

O'Brien, T. (2004). *Writing in a foreign language: Teaching and learning. Language Teaching, 37*(1), 1-28. doi:10.1017/S0261444804002113

O'Malley, J., Chamot, A., Stewner-Manzanares, G., Kupper, L., & Russo, R. (1985) Learning strategies used by beginning and intermediate ESL students. *Language Learning, 35*(1), 21–46.

OneNote (Version 16.0.7127.1026, 2013). Microsoft Home and Student (Software).

Ong, J. (2014). How do planning time and task conditions affect metacognitive processes of L2 writers? *Journal of Second Language Writing, 23,* 17–30.

Ortega, L. (2009). Studying writing across EFL contexts: Looking back and moving forward. In R. Manchón (Ed.), *Writing in Foreign Language contexts: Learning, Teaching and Research* (pp. 232 - 255). Bristol: Multilingual Matters.

Ortega, L. (2012). Epilogue: Exploring L2 writing–SLA interfaces. *Journal of Second Language Writing, 21,* 404-415. doi:10.1016/j.jslw.2012.09.002

Oxford, R. (2001). Language learning strategies. In R. Carter & D. Nunan, (Eds.), *The Cambridge guide to TESOL* (pp. 166–185). Cambridge, UK: Cambridge University Press.

Oxford, R. (2011). *Teaching and researching language learning strategies.* Harlow, England: Pearson Longman.

Padilla, A., Aninao, J., & Sung, H. (1996). Development and implementation of student portfolios in foreign language programs. *Foreign Language Annals, 29*(3), 430-438. doi:10.1111/j.1944-9720.1996.tb01254.x

Paesani, K. (2006). Exercices de style: Developing multiple competencies through writing portfolio. *Foreign Language Annals, 39*(4), 618-638.

Pica, T. (1984). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition, 6*(1), 69-78. doi:10.1017/S0272263100000309

Picot, B. (1988). *Administering for excellence: effective administration in education: Report of the taskforce to review education administration.* Wellington, N.Z: MOE

Platt, E. & Brooks, F. (1994). The "Acquisition-Rich Environment" revisited. T*he Modern Language Journal, 78*(4), 497-511. doi: 10.1111/j.1540-4781.1994.tb02067.x

Polio. C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21,* 375-389.

Polio, C., & Friedman, D. (2017). *Understanding, evaluating, and conducting second language writing research.* Second Language Acquisition Series. NY: Routledge.

Porto, M. (2007). Learning diaries in the English as a foreign language classroom: A tool for accessing learners' perceptions of lessons and developing learner autonomy and reflection. *Foreign Language Annals, 40*(4), 672-696. doi:10.1111/j.1944-9720.2007.tb02887.x

Ramesh, R. (2009). Metacognitive and cognitive strategies for enhancing second language acquisition. In R. Ramesh (Ed.), *Innovative Exploration in English Language Teaching.* Karaikudi: Vinsi Publishers.

Ransdell, S., & Barbier, M. (2002). An introduction to new directions for research in L2 writing. In S. Ransdell & M. Barbier (Eds.), *Studies in writing 11: New directions for research in L2 writing* (pp. 1-10). Dordrecht, NLD: Kluwer Academic Publishers.

Reichelt, M. (1999). Toward a more comprehensive view of L2 writing: Foreign language writing in the U.S. *Journal of Second Language Writing, 8*(2), 181-204. doi:10.1016/S1060-3743(99)80127-1

Reichelt, M. (2011). Foreign Language Writing: An overview. In T. Cimasko & M. Reichelt (Eds.), *Foreign language writing instruction: Principles and practices* (pp.3-21). Anderson, SC: Parlor Press.

Richards, J. (2006). *Communicative Language Teaching today.* New York: NY: Cambridge University Press.

Renou, J., 2001. An examination of the relationship between metalinguistic awareness and second language proficiency of adult learners of French. *Language Awareness 10,* 248–267.

Roberts, M., & Griffiths, C. (2008). Errors correction and good language learners. In C. Griffiths (Ed.). *Lessons from Good Language Learners.* Cambridge: Cambridge University Press, p. 282–293.

Roca de Larios, J., Manchón, R., Murphy, L., & Marin, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing, 17*, 30-47. doi:10.1016/j.jslw.2007.08.005

Roca De Larios, J., & Murphy, L. (2001). Some steps towards a socio-cognitive interpretation of second language composition processes. *International Journal of English Studies, 1*(2), 25-45.

Roman, D., Wientjes G., Thompson, K., & Hakuta, K. (2009). Wordsift: An interactive web-based vocabulary tool designed to enhance academic literacy. AccELLerate 1(4), 14-16.

Rubin, J. (1975). *W*hat the 'good language learner' can teach us. *TESOL Quarterly, 9*(1), 41-45.

Rubin, J. (2003). Diary writing as a process: simple, useful, powerful. *Guidelines, 25*(2), 10-14.

Rubin, J. (2008). Reflections. In Griffiths, C. (Ed.), *Lessons from Good Language Learners* (pp. 10-16). Cambridge: Cambridge University Press.

Ruiz-Funes, M. (2015). Exploring the potential of second/foreign language writing for language learning: The effects of task factors and learner variables. *Journal of Second Language Writing, 28,* 1-19.

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition, 29,* 67-100.

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing, 9*(3), 259-291. doi:10.1016/S1060-3743(00)00028-X

Scardamalia, M., Bereiter, C., & Steinbach, R. (1984). Teachability of reflective processes in written composition. *Cognitive Science, 8*(2), 173-190. doi:10.1207/s15516709cog0802_4

Schunk, D., & Usher, E. (2013). Barry J. Zimmerman's theory of self-regulated learning. In H. Bembernutty, T. Cleary, & A. Kitsantas. (Eds.), *Applications of self-regulated learning across diverse disciplines: A tribute to Barry J. Zimmerman* (pp. 1-28). Charlotte NC: Information Age Publishing, Inc.

Schunk, D., & Zimmerman, B. (2009). *Motivation and self-regulated learning: Theory, research and applications*. NY: Routledge.

Schoonen, R., Snellings, P., Stevenson, M., & Van Gelderen, A. (2009). Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R. M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 77–101). Bristol, UK: Multilingual Matters.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching, 10*(3), 209-231.

Sewell, K. (2011, Sept. 14). Forward to the New Zealand curriculum. Available: http://nzcurriculum.tki.org.nz/Curriculum-documents/The-New-Zealand-Curriculum/Foreword

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning, 64*(1), 103-131.

Simard, D., French, L., & Fortier, V. (2007). Elicited metalinguistic reflection and second language learning: Is there a link? *System, 35,* 509–522

Slavkov, N. (2015). Sociocultural theory, L2 writing process, and Google Drive: Strange bedfellows? T*ESL Canada Journal/Revue TESL du Canada, 32(*2), 80-94.

Smith, H. (2007). The Social and Private Worlds of Speech: Speech for Inter- and Intramental Activity. *The Modern Language Journal, 91*(3), pp. 341-356.

SPSS Statistics (Version 23, 2015). IBM (Software).

Stern, H. (1975). What can we learn from the good language learner? *Canadian Modern Language Review, 31*, 304–318.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), *Input and second language acquisition* (pp. 235-252). Rowley, Mass: Newbury House.

Swain, M. (2006) Languaging, agency and collaboration in advanced second language learning. In H. Byrnes (Ed.), *Advanced Language Learning: The contributions of Halliday and Vygotsky.* London: Continuum.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*(3), 371-91. doi:10.1093/applin/16.3.371

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal, 82,* 320-337.

Swain, M., & Lapkin, S. (2000). Task-based second language learning: the uses of the first language. *Language Teaching Research, 4*(3), 251-274.

Te Kete Ipurangi (2012). *The Online Curriculum Guide.* Available: http://seniorsecondary.tki.org.nz/Learning-languages/What-s-new-or-different#communication

Teng, S. & Zhang, L. (2016). A questionnaire-based validation of multidimensional models of self-regulated learning strategies. *The Modern Language Journal, 100*(3), 674-701.

Torras, R., & Celaya, L. (2001). Age-related differences in the development of written production. An empirical study of EFL school learners. *International Journal of English Studies, 1*(2), 103-126.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327-369.

Turnbull, M., & Dailey-O'Cain, J. (2009). *First language Use in Second and Foreign Language Learning.* UK: Multilingual Matters.

Ushioda, E. (2009). A person-on-context relational view of mergent motivation, self, and identity. In Z. Dörnyei. & E. Ushioda. (Eds.), *Motivation, Identity and the L2 Self* (pp. 215-28). Bristol: Multilingual Matters.

Van Beuningen, C., De Jong, N., & Kuiken, F. (2012). Evidence of the effectiveness of comprehensive error correction in second language writing. *Language Learning 62*(1),1-41.

Vygotsky, L. (1978). In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in society: The development of higher psychological processes* (A. Luria Trans.). Cambridge, Massachusetts: Harvard University Press.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching, 31,* 57-71.

Wenden, A. (1999). An introduction to metacognitive knowledge and beliefs in language learning: Beyond the basics. *System, 27,* 435-441.

Werner, H. (1956). Microgenesis and aphasia. *Journal of Abnormal Social Psychology, 52*, 347-353.

Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing, 21,* 364–374. doi.org/10.1016/j.jslw.2012.09.005

Williams, M., & Burden, R. L. (1997). *Psychology for language teachers.* Cambridge UK: Cambridge University Press.

Wolff, D. (2000). Second language writing: A few remarks on psycholinguistic and instructional issues. *Learning and Instruction, 10*(1), 107-112. doi:10.1016/S0959-4752(99)00021-3

Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity.* Honolulu: Second Language Teaching and Curriculum Center, University of Hawaii at Manoa.

Word (Version 14.0.7173.5000, 2010). Microsoft Home and Student (Software).

Wordsift (Version 2). Available: www.wordsift.org

Wylie, C., & Bonne, L. (2016, May). *Secondary schools in 2015: Findings from the NZCER national survey.* Wellington: NZCER. Available: http://www.nzcer.org.nz/research/publications/secondary-schools-2015

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication, 50*(3), 483-503.

Zhang, L. (2008). A Comprehensive Review of Studies on Second Language Writing. *HKBU Papers in Applied Language Studies, 12,* 89-123.

Zhang, L. (2013). Language learning strategies (LLS) and Strategy-based instruction (SBI): What can we do in ESOL? (Presentation). Faculty of Education, University of Auckland, NZ.

Zimmerman, B. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), p. 3-17.

Zimmerman, B. (2009). Goal setting: A key proactive source of academic self-regulation. In D. Schunk & B Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research and application* (pp. 267-296). NY: Routledge.

Zimmerman, B. (2013). Foreword. In H. Bembernutty, T. Cleary, & A. Kitsantas (Eds.), *Applications of self-regulated learning across diverse disciplines: A tribute to Barry J. Zimmerman.* Charlotte NC: Information Age Publishing.

Zimmerman, B., & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal, 23,* p. 614-628.

Zimmerman, B., & Schunk, D. (2009). Motivation: An essential dimension of self-regulated learning. In *Motivation and self-regulated learning: Theory, research and application* (pp. 1-30). D. Schunk & B Zimmerman (Eds.). NY: Routledge.

APPENDICES

Appendix A: NCEA Level 1 Writing Standard (abbreviated)

This achievement standard involves writing a variety of text types to communicate in French on areas of most immediate relevance.

Achievement Criteria

Achievement	Achievement with Merit	Achievement with Excellence
• Write a variety of text types in French on areas of most immediate relevance.	• Write a variety of text types in convincing French on areas of most immediate relevance.	• Write a variety of text types in effective French on areas of most immediate relevance.

Explanatory Notes

1   This achievement standard is derived from the Learning Languages Communication Strand Curriculum Level 6 of *The New Zealand Curriculum*, Learning Media, Ministry of Education, 2007, and is related to the material in the *Teaching and Learning Guide for Languages*, Ministry of Education, 2010, at http://seniorsecondary.tki.org.New Zealand.

2   Curriculum Level 6 includes the communication skills, language and cultural knowledge needed to write a variety of texts on areas of most immediate relevance.

3   Definitions
Write refers to composing and organising text in a linguistically and culturally appropriate format and style, and may include:
   • drafting
   • reworking over a period of time.

Write on areas of most immediate relevance refers to using language, related to basic personal information and past, present, and/or future experiences, in order to express personal information, ideas and opinions in culturally appropriate written French.

Communication is achieved overall, despite inconsistencies, such as:
   • format
   • spelling
   • lexical choice
   • level of formality
   • language conventions
   • language features.

Convincing means that there is development of the information, ideas and opinions which is generally credible and connected. The writer selects and uses a range of language and language features that are fit for purpose and audience. Communication is not significantly hindered by inconsistencies.

Effective means that there is development of the information, ideas and opinions which is controlled and integrated. The writer capably selects and successfully uses language and language features that are fit for purpose and audience. Communication is not hindered by inconsistencies.

Variety refers to texts selected from a range of different text types, which have been created for different purposes. At all times the quality of the texts in the selection, considered as a whole, is more important than length.

4   Conditions of Assessment related to this achievement standard can be found at www.tki.org.New Zealand/e/community/ncea/conditions-assessment.php.

Appendix B: NCEA Level 1 French Assessment Schedule

Evidence/Judgments for Achievement	Evidence/Judgments for Achievement with Merit	Evidence/Judgments for Achievement with Excellence
The student writes a variety of text types in French on areas of most immediate relevance.	The student writes a variety of text types in convincing French on areas of most immediate relevance.	The student writes a variety of text types in effective French on areas of most immediate relevance.
The student provides a collection of at least two examples of written texts.	The student provides a collection of at least two examples of written texts.	The student provides a collection of at least two examples of written texts.
The student texts are of different text types and are for genuine purposes.	The student texts are of different text types and are for genuine purposes.	The student texts are of different text types and are for genuine purposes.
The texts provide a total of about 300 words.	The texts provide a total of about 300 words.	The texts provide a total of about 300 words.
The texts communicate basic personal information, ideas and opinions.	The texts communicate basic personal information, ideas and opinions.	The texts communicate basic personal information, ideas and opinions.
Over the texts there is evidence of communicating beyond the immediate context, for example, about the past and/or future.	Over the texts there is evidence of communicating beyond the immediate context, for example, about the past and/or future.	Over the texts there is evidence of communicating beyond the immediate context, for example, about the past and/or future.
The student achieves communication overall, despite inconsistencies in aspects such as format, spelling, lexical choice, level of formality, language conventions, or language features.	The texts show development of the information, ideas and opinions that is generally credible and connected.	The texts show development of the information, ideas and opinions that is controlled and integrated.
	The student selects and uses a range of language and language features that are fit for the purpose and audience of the tasks.	The student capably selects and successfully uses a range of language and language features that are fit for the purpose and audience of the tasks.
	Inconsistencies do not significantly hinder communication.	Inconsistencies do not hinder communication.
For example: *Ma ville se trouve dans le sud de la Nouvelle-Zélande. Il y a beaucoup de choses à faire. J'adore ma ville. Pendant les vacances je suis allé à Auckland.* The examples above are indicative samples only.	For example: *Ma ville se trouve dans le sud de la Nouvelle-Zélande. C'est bien parce qu'il y a beaucoup de choses à faire. Il faut visiter le musée. Pendant les vacances de Pâques je suis allé à Auckland pour visiter mes cousins et nous avons fait beaucoup de choses.* The examples above are indicative samples only.	For example: *Ma ville c'est la capitale, Wellington, et elle se trouve dans l'île du nord en Nouvelle-Zélande. C'est très amusant ici parce qu'il y a toujours beaucoup de choses à faire. Si vous visitez ma ville il faut surtout visiter le célèbre musée de Te Papa. Pendant les vacances de Pâques je suis allé à Auckland et j'ai fait beaucoup de choses mais à mon avis Auckland est moins intéressant que Wellington.* The examples above are indicative samples only.

Appendix C: Writing Task 1, Unit 1

En Échange

Portfolio Writing Task: Personal Profile Draft 1 (test)

This task arises from the genuine context of an online exchange. Your purpose is to introduce yourself in a profile that could be posted online and to indicate what you would like in an exchange partner.

You could include:
- Personal details – name. age, birthday, brothers and sisters, where you live, pets etc
- Daily routine
- Likes and dislikes with reasons
- Friends and family relationships
- What you did last weekend and what you intend to do next weekend
- Your ideal correspondant/e

Conditions of Assessment (for the portfolio as a whole)

- There is no word limit for the portfolio but it is recommended that, across the three texts, you write a total of about 300 words. Quality is however more important than quantity.
- This first draft is written under test conditions and will be used to guarantee the work you produce is your own. In one week's time, you will be given general feedback on your first draft and feedback on the common errors of the whole class. You then have one week to rework your writing before you hand in a second draft, that is, on _____. Keep all your drafts. However, only your final version will be assessed. Make certain you label all versions clearly to avoid loss and confusion.
- You may use resources such as word lists, dictionaries, text books, native speakers and grammar notes to help you when you are drafting and revising you work but you may not ask someone else to proofread or correct your writing.
- All work must be your own work. Extracts from external sources should not be included without acknowledging the sources. Any extracts from external sources will not be considered in the final achievement judgement. Language from the language samples in the assessment schedule may not be used unless it is significantly reworked. For the school's assessment policy, refer to the authenticity statement below.
- The final three texts selected are considered as a whole for one grade allocation.

Appendix D: Writing Task 2, Unit 2

En Échange

Portfolio Writing Task: Blog (Extended Online Postcard)

In one week's time, on _____, draft 1 of this blog (extended online postcard) task will be written in a 30 minute test. You will have no resources other than the supervising teacher. You have one week to prepare for it. You are **either:**

- a New Zealander hosting a French-speaking exchange student and taking time out to show him/her places of interest in New Zealand **or**
- a New Zealander on exchange in a French-speaking country where your host family is taking time out to show you places of interest in their part of the world.

You are to write a blog entry (extended online postcard) telling family and friends what you have been up to and what your plans are for the next week. You could include:

- Where you are both staying and how you are travelling
- What the weather has been like
- The places you have visited and the activities you have taken part in
- The best and the worst parts of the trip so far and why
- An incident that has occurred (e.g. a loss, an accident, a disagreement, an opportune meeting)
- What your plans are for the rest of the trip

Conditions of Assessment (for the portfolio as a whole)

- There is no word limit for the portfolio as a whole but it is recommended that, across the three texts, you write about 300 words. Remember that quality is more important than quantity.
- The first draft will be written under test conditions and will be used to guarantee the work you produce is your own and nobody else's. One week later, you will be given coded feedback on your first draft and teaching on the common errors of the whole class. You then have one further week to rework your writing.
- You will hand in your second draft on

  _____.

- Keep all your drafts. However, only your final version will be assessed. Make certain you label all versions clearly to avoid loss and confusion. Your teacher will tell you how and where to store your drafts:

  _____.

- When you are redrafting and improving your work, you may use resources such as word lists, dictionaries, text books, native speakers and grammar notes to help you but you must not ask anyone to proofread or correct your writing.
- This is because all work must be your own work. Extracts from external sources should not be included without acknowledging the sources. Any extracts from external sources will not be considered in the final achievement judgement. Language from the language samples in the assessment schedule may not be used unless it is significantly reworked.
- Please ask your classroom teacher about your school's authenticity statement.

The final three texts that you select for inclusion in your portfolio will be considered as one whole and be awarded one grade. Any grades given on drafts are indicators only. They will help you to make further improvements.

Appendix E: Reflection Template

En Échange

You have just written and received feedback on the first draft of an online profile in French. Now you must prepare and write a second draft for homework. Answer the following questions as honestly, thoughtfully and fully as you can. Use the notes you took after writing the first draft to help you with Section A.

Section A: You may refer to the notes you made on the day of the first draft test:

1. Before the test: How did you prepare for the first draft? (Length of time, strategies you used, the process you went through.) Are you happy with your preparation? What do you want to remember/change for the next time?

2. During the test: As you were writing in French, what difficulties did you have? What strategies did you use to overcome the difficulties? Give specific examples. What went well? What do you want to remember for next time?

Section B: On the day you receive your feedback

3. Feedback on Draft 1: Tally up the codes in the coded feedback you have just received. What are your three most frequent errors? Are you surprised? Why? Why not? What do you intend to do about it before you write your second draft? Did the codes prove helpful? Why? Why not? What about the teaching on common errors? Was that helpful? Will it help you improve your writing in the second draft?

4. Feed forward for Draft 2: Look at the teacher's comments and the NCEA grade indicator. Are they helpful? Have they helped you see where you can improve your second draft?

5. Further comments: Any further thoughts on the processes and strategies you used for this first draft of writing assessment and for the second draft you are about to write?

Note: Second draft due on _____.

Appendix F: Coding Schedule for Corrective Error Feedback

Code	Meaning	Hints	Practice/Monitor
+	**You have redundant words**	• Do you need all the words in this sentence? • Have you repeated yourself?	Backtranslate
∧	**You have omitted a word/words**	• Have you checked that all the words you need are present? • What is missing in this sentence?	Backtranslate
A	**Agreements**	• Have you checked the genders of your nouns? • Do your adjectives agree with those nouns in number and gender? • Do your articles match your nouns in number and gender? • Have you used an adjective that does not agree?	Online exercises: https://www.laits.utexas.edu/tex/gr/index.html
HC	**Hindered Communication**	• Can you rewrite this section to make it more meaningful? • Is there a section that needs to be simplified so communication is achieved? • Can you clarify your thoughts and language on this line? • Have you said what you meant to say? • Does this sentence make sense to you?	Backtranslate
L	**Articles, particles, conjunctions and connecting and linking words of all kinds**	• Have you chosen the right articles and particles? • Have you chosen the right prepositions, preceding direct and indirect objects and negatives and placed them in the right position? • Have you used the best connecting and linking words?	Online exercises: https://www.laits.utexas.edu/tex/gr/index.html
R	**Register**	• Is your choice of language appropriate for the text type, audience and purpose? • Have you used the most appropriate word, *Tu* or *vous*? • Have you been consistent in its use?	Tu/vous check Tone check
Sp	**Spelling**	• Have you checked your spelling? • Have you checked your accents? • Do you have capital letters where needed or not needed? • Are your hyphens, full stops, commas and gaps in the right place?	Language Perfect English to French
U	**Idioms and formulaic expressions versus franglais (Usage)**	• Have you misquoted a formulaic expression? • Is this the way French people say this? • Are you thinking in English instead of French? • Have you guessed and produced a word that does not exist in French? • Have you translated from English instead of using a French idiom? • Have you used English words or structures instead of French ones? • What about the difference between French and English word order?	Wordreference.com both ways. Follow up with vocab learning strategies.
V	**Verb**	• Have you chosen the correct tense for your verbs? • Have you checked that your subjects match your verbs? • Are these regular or irregular verbs? Check the stems! • If you are using the passé composé, do you have the right auxiliary? Avoir or être? • If you have an 'être' verb, does the subject agree with the past participle?	Online exercises: https://www.laits.utexas.edu/tex/gr/index.html
WO	**Word Order**	• Which adjectives go after the noun? • Where do negative particles go? • Where do adverbs go? • Where do object pronouns go?	Online exercises: https://www.laits.utexas.edu/tex/gr/index.html
WW	**Wrong Word**	• Do you have the right word for the meaning you want to convey? • Is there a better word to use here? • Have you done a two-way dictionary check? • Have you guessed and produced a word rather than looking it up?	Wordreference.com both ways. Follow up with vocab learning strategies.

Appendix G: Situated Version of East's (2008) Scoring Rubric

	Content coverage and flow of ideas	Knowledge of vocabulary and its functional and idiomatic use	Grammatical variety and sentence-level competence	Spelling & punctuation, including accents	Intercultural understanding (register, awareness of reader and text-type)
7	• covers all 7points in the instructions • includes opinions and reasons • very logical sequencing • writing 'flows'	• very effective choice & usage of words / idioms /functions • there is solid mastery of word form • the meaning is not obscured	•very few errors and of a very minor nature	• very few errors of spelling, punctuation, accents • the meaning is not obscured	• very aware of French reader e.g. Task 1: compares NZ and France and requests and describes penpal. Task 2: rich descriptions of appropriate tourist sites and activities •very appropriate opening and closing
6	•covers at least 6 points in the instructions •includes opinions and reasons • logical sequencing • writing generally ''flows''	•effective choice & usage of words / idioms / functions •there is mastery of word form • the meaning is not obscured	• a few errors but generally of a minor nature	• few errors of spelling, punctuation, accents • the meaning is not obscured	• generally aware of French reader e.g. Task 1: describes / gives opinion about NZ/town or 'translates' e.g. en seconde and requests and describes penpal. Task 2: descriptions of appropriate tourist sites / activities • good opening and closing
5	• covers at least 5 points in the instructions) • includes opinions and reasons • mainly logical but some incomplete sequencing • writing is a little bit 'choppy'	• there are occasional errors in choice & usage of words / idioms / functions • the meaning is very rarely obscured	•simple constructions used accurately •minor problems with complex constructions •several errors • the meaning is very rarely obscured	• occasional errors of spelling, punctuation, accents • the meaning is very rarely obscured	• some awareness of French reader e.g. Task 1:Refers to NZ e.g. j'habite en N-Z and requests a penpal Task 2: Describes at least 1 appropriate tourist site / activity. • satisfactory opening and closing
4	• covers at least 4 points in the instructions) • includes opinions • generally logical but several instances of incomplete sequencing • writing is somewhat 'choppy'	• there are more frequent errors in choice & usage of words / idioms / functions • the meaning is seldom obscured	•simple constructions are mostly used accurately • more noticeable problems with complex constructions •several errors • the meaning is seldom obscured	• more frequent errors of spelling, punctuation, accents, • the meaning is seldom obscured	• indicates some awareness of reader e.g. Task 1: greets and/or may mentions NZ and/or requests a penpal. Task 2: Refers to at least 1 appropriate tourist site / activity •attempts opening and closing
3	• covers at least 3 points in the instructions) • includes opinions • lacks logical sequencing & development • writing is quite 'choppy'	• there are frequent errors in choice & usage of words / idioms / functions • the meaning is somewhat confused or obscured	• major problems with both simple & complex constructions •frequent errors • the meaning is somewhat confused or obscured	• frequent errors of spelling, punctuation, accents • the meaning is somewhat confused or obscured	• little awareness of reader e.g.Task 1: omits greetings and request for penpal and reference to NZ. Task 2: reference to place or activity of minor interest to tourists • attempts opening or closing
2	• covers at least 2 points in the instructions) • very minimal logical sequencing & development • writing is very 'choppy'	• there are very frequent errors in choice & usage of words / idioms / functions • communicates, but the meaning is confused or obscured	• some very minor mastery of sentence construction rules •mostly dominated by error • communicates, but the meaning is confused or obscured	• very frequent errors of spelling, punctuation, accents • the meaning is confused or obscured	• no awareness of French reader e.g. Task 1: no greetings, no penpal request and information of very minor interest. Task 2: Reference to site or activity of very little interest to tourists • little attempt at opening or closing
1	• covers only 1 point in the instructions • virtually no logical sequencing & development • writing is extremely 'choppy'	• there is little knowledge of vocabulary, idioms, functions • the meaning is very confused or obscured	•virtually no mastery of sentence construction rules •almost exclusively dominated by error • the meaning is very confused or obscured	• demonstrates very limited mastery of conventions • dominated by errors of spelling, punctuation, accents, etc • the meaning is very confused or obscured	• barely communicates personal information, ideas & opinions •there is very inadequate language for the text-type • no attempt at opening or closing
0	• no rewardable response				

Appendix H: Rubric of Aljaafreh and Lantolf's (1994) Regulatory Scale

Microgenetic Growth: Five Levels of Transition

Developmental Stage		Reduction in frequency / eradication of error		Reduction in reliance on teacher	
Stage 1 Other-regulation in which the learner relies on another to perform OR Intermental	**Level 1**	The learner is not able to notice the error. *Many errors.*	The learner is not able to correct the error even with intervention from the teacher. *Most errors remain in 2nd draft.*	The learner does not have a sufficient basis from which to interpret the teacher's moves to provide help, and probably has no awareness that there is even a problem.	The teacher must assume full responsibility for correcting error. Rather than providing corrective help, the teacher's task is to bring the target form into focus, and begin the process of co-constructing the ZPD with the learner.
	Level 2	The learner is able to notice the error. *Many errors*	The learner cannot correct the error, even with intervention. *25% of errors corrected.*	The learner must still rely heavily on the teacher.	Because there is some degree of development, an opening is provided for the teacher and the learner to begin negotiating the feedback process and for the learner to begin to progress toward self-regulation. Help tends to be explicit.
	Level 3	The learner is able to notice error under other-regulation. *Fewer errors.*	The learner is able to correct error under other-regulation. *50% of errors corrected.*	The learner understands the teacher's intervention and is able to react to the feedback offered.	The levels of help needed to correct the error move toward the strategic, implicit end of the regulatory scale.
Stage 2 Partial self-regulation OR Partially intramental	**Level 4**	The learner notices error with minimal or no obvious feedback from the teacher. *Fewer errors.*	The learner corrects error with minimal or no obvious feedback from the teacher. *75% of errors corrected.*	The learner begins to assume full responsibility for error correction. The learner may even reject feedback from the teacher when it is unsolicited (e.g., "Let me see if I can do it alone").	Development has not yet become fully intramental or automatised, since the learner often produces the target form incorrectly and may still need the teacher to confirm the adequacy of the correction.
Stage 3 Full self-regulation OR Fully intramental	**Level 5**	Whenever aberrant performance arises, noticing errors does not require intervention from someone else. *Minor slips.*	Correcting errors does not require intervention from someone else. Mistakes emanate from legitimate slips of the tongue or pen rather than from incomplete learning. *Evidence of self-correction.*	The learner becomes more consistent in using the target structure correctly in all contexts. In most cases, the individual's use of the correct target form is automatised.	The individual is fully self-regulated, including corrective behaviour, is completely self-generated and automatised.

Note: Words in italics have been added. All other descriptors come directly from Aljaafreh and Lantolf (1994, p. 470-471).

Appendix J: Compilation of Strategies

Research stage	Writing-to-learn-language processes*	Strategies employed to achieve writing-to-learn-language processes*
Stage 1: Planning for the writing test	Advance preparation, retrieving from long-term memory phrases that match the task requirements Managing the writing-to-learn-language process Evaluating and overcoming knowledge gaps in communicating, considering how much distance there is between known language and language needed to respond to the task	Getting in the right frame of mind [G] (generating and maintaining motivation [O]) Activating emotions, beliefs and attitudes [O] Giving self-consequences [Z] Understanding the task [G] Goal setting [Z] Planning [O] [Z] Implementing plans [O] Orchestrating strategies [O] Organising [O] [Z] Obtaining and using resources [O] (information seeking [Z]) Environmental structuring [Z] Activating prior knowledge [O] Reviewing notes, books, tests [Z] Mind mapping [G] Brainstorming [G] Thinking in L1 [G] Thinking in L2 [G] Rehearsing [Z] Memorising [Z] Dealing with sociocultural contexts and identities [O] Evaluating the gap [G] Overcoming knowledge gaps [O] Seeking social assistance from peers, teacher, or other adults [Z] (interacting to learn and communicate [O])
Stage 2: Composing the first draft during the test	Formulating (going from an idea to the first attempt at drafting), applying L2 knowledge, and monitoring, checking for mistakes Noticing-triggering and hypothesis-testing	Recombining a set phrase with another set phrase [G] Restructuring or changing a set phrase [G] Translating [G] Avoiding [G] Paying attention [O] Monitoring [G] [O] (self-monitoring [Z]) Visual monitoring [G] (using senses to understand, remember [O]) Auditory monitoring [G] (using senses to understand, remember [O]) Back-translating [G] Prompting a specific monitoring of vocabulary or grammar [G] Common sense monitoring [G] Reasoning [O]
Stage 3: Reviewing and Writing the second draft	Engaging critically with feedback (indirect error codes, teaching on common errors, general comments, NCEA grade indicator) to improve L2 writing and strategy use Metalinguistic reflection	Reacting to feedback [G] Self-evaluating [Z] (evaluating [O]) Completing a clear practical task after receiving feedback on error [G] Changing, adjusting, transforming [Z] Record keeping [O] Conceptualising with details [O] Conceptualising broadly [O] Going beyond the immediate data [O]

*Despite the tabular divisions above, L2 writing is a recursive rather than linear exercise so, while some process-strategy combinations are more likely than others, processes and strategies are used in any order and in any combination.
[G] Graham and Macaro (2007)
[O] Oxford (2011)
[Z] Zimmerman (1990)

Appendix K: Adapted Version of the SOLO Taxonomy

Level	Description
1. Incompetence:	Learner misses the point. (Thinking may indicate the learner is unaware, passive, or unengaged.)
2. Unistructural:	Learner can name, identify single relevant aspects or follow simple procedures. (Thinking may be vague, fuzzy or general, show superficial awareness, offer some information but not enough, or indicate reliance on others.)
3. Multistructural:	Learner can combine, describe, enumerate, list several relevant independent aspects or perform serial skills. (Thinking may be specific and purposeful but lack in detail and full understanding.)
4. Relational:	Learner can integrate ideas into a structure, analyse, apply, argue, compare/contrast, criticise, explain causes, relate, or justify. (Thinking may be specific, purposeful, detailed, and indicate full understanding.)
5. Extended Abstract:	Learner can generalise ideas to new domains, can create, formulate, generate, hypothesise, reflect, and theorise. (Thinking may be specific, purposeful, detailed, and indicate full understanding that is able to be transferred to new situations.)

Appendix L: Examples of L2 Writing and L1 Reflections

Achieved Participant's L2 Writing and L1 Reflection (Task 2):

Cher Harris

Normalement, quand je suis en vacances, je visite mes copains à Hamilton et fais du shopping et vas au cinéma, mais cette fois, ma famille et moi avons décidé de sortir à ville donc tu peux vois notre pays. Néo Zéalande pour est ce-que.

Je suis allé aux plages dans l'île de nord mais quand tu es rentré, nous penvons penvons visitons tous les places et faissons du planche à voile! Je ne fait rien les sports d'aqua alors il étais interressant pour nos.

Je vraiment veux tu arriver tôt parce que les vacances à est en être assez triste. Mon frère est monté à l'escalier donc il a avair cassé son pied. C'était Ce n'était pas tres cool! Mais quand tu en arrive lui, sa jambe son pied étas bonne aller!

Tous les monds restais dans l'hôtel d'Auckland - City Sky Tower, et tu et moi iras nos personalle salle à l'hôtel! Je ne peux pas

De plus, les meuillers magasines pour les vêtements sont un deux minutes à la l'hôtel si nous marches marchons!

Je suis très content et je n'attend pas pour tui!

Ammicalement,

1. Language Perfect - the proper sentence structures help not the word remembering. I didn't spend a lot of time preparing for this assignment unfortunately, or at least not as much as I wanted to. I did reread the Studio Trois Livre and filled in whenever I had no homework in the Studio Trois exercise book we were given.
2. I had struggles in making sure my sentence structures were correct. To overcome this, I just write down the sentence, reread it, remembered the rules we are taught in class and then rewrite it as best as I can. This helps me in ensuring what I write feels right to what I write in class. I had a couple mind blanks for what certain words were in French but know them in Bulgarian and English. To overcome this, I tried to see if there was a root word within both translations and tried to fit it in a sentence, remembering what the french version would be using that root word.
3. …
4. the NCEA grade indicator has helped me understand what I definitely need to work on at home in preparation for my second draft. The teacher's comment helped me understand that my grade was very low and that my sentences were confusing but did not state how else I could try get my grade up to a strong achieved or merit. However as she mentioned the sentences were hard to understand, I knew to just refer to the codes and hope that the codes, once fixed, would get my level of grade up.
5. Language Perfect - the sentence structures and verb usage. Language Perfect - the word remembrances (NCEA List). French Year 10 Revision. Studio Trois Exercise Book. Cahier Gris Grammare Structeur Book. Practicing writing at home, using Studior Trois Livre to help.

Merit Participant's L2 Writing and L1 Reflection (Task 2):

Bonjour chères lecteurs,
malheureusement mon corres Zacharie doit retourner en France demain.
Je suis triste, mais on s'est bien amusé. Nous restons dans l'hôtel
<< Waterfront >> au Quay Street, parce que c'est plus rapide d'aller
en choses au hôtel qu'au chez moi. Nous avons prendre le bus à travers,
parce que ce n'est pas cher, et c'est plus facile et rapide que aller en voiture.
Nous avons vu beaucoup de choses. Premièrement, nous sommes allés au
<< Sky Tower >> C'était fantastique! Les vues étaient magnifiques, et il
a fait beau aussi. Nous sommes allés aussi en << Kelly Tarltons >>
mais j'ai oublié mon argent, et on ne peut peut pas entrer. Alors, on
doit retourner à l'hôtel pour mon argent, et puis, nous sommes encore
aller au << Kelly Tarltons >> Le << Sky Tower >> était magnifique, mais la
chose le plus marquant était les jardins botaniques! Ils ont beaucoup
de belles fleurs, et, en plus, le "Titan Arum" est en fleur depuis deux
jours. C'est une chose que s'est passé seulement une fois par décennie!
Demain, nous allons au restaurant << Euro >> l'après-midi, pour manger
du poisson le plus savoureux à Auckland, et puis, Zacharie doit retourner
en France. J'espère que Zacharie peut retourner en Nouvelle-Zélande
parfois.

Amicalement,

1. The night before, I spent maybe around half an hour to an hour writing a draft copy, then at school the next morning I wrote it out on a piece of refill to help memorise it. For next time, I would keep the same strategy, because it worked quite well for memorising the bulk of the content. Most of the errors
I made were there in my written draft, so it was not an issue of messing up on the day.
2. I didn't have many difficulties writing in French, but as per my first draft, I made a lot of errors, although this is a result of my pre-prepared draft not being very accurate and correct in the first place. Next time, I should put more time into preparing a draft to use.
3. My number 1 error was verbs, because I neglected to use past tense, which I am not surprised about since I never was an expert on verb conjugation. Articles were number 2, and spelling number 3. I am a little surprised about articles, but not so much about spelling, because I was also not the best at it. The codes are useful because they help to identify where the error lies, e.g. if it is a verb error, you will know to look at verbs. The general comment also aids the codes, because it helps point you in the right direction for fixing your errors.
4. As in (3), the comments proved useful in addition to the codes, as they pointed me in the right direction to fix the errors, specifically for verbs, because it mentioned the specific faults in my writing. The NCEA grade indicator is also useful, because it helps to give an idea of what you should be looking for to add into your draft to improve your grade.

Excellence Participant's L2 Writing and L1 Reflection (Task 2):

Je suis trop fatiguée maintenant, mais aussi très content car ma étudiante étrangère Rosie et moi, on est allées aujourd'hui aux jardins botaniques pour voir toutes les belles fleurs. C'était fatigant parce qu'il a commencé à pleuvoir pendant notre pique-nique près du jardin des roses à treize heures et alors on a dû prendre toutes les choses et courir au café pour manger. Pourtant, les jardins botaniques sont plus grands et très magnifiques! On s'est bien amusées ensemble.

Rosie et moi, nous restons chez ma tante à la campagne pour une semaine, comme Rosie adore les fermes et ma tante Megan a beaucoup de vaches et de moutons ici. Il fait d'habitude assez froid comme c'est hiver en ce moment, mais nous espérons qu'il va faire beau demain, alors nous pourrons aller à la Skytower et prendre des photos. Pour aller en ville, on doit prendre le bus, parce que c'est moins cher et c'est l'option facile, oui?

On veut monter en haut de la Skytower et on veut aussi y manger, même que c'est assez cher. Rosie m'a dit que la tour d'Eiffel est moins grande que la Skytower, et elle a voulu y visiter depuis qu'elle est arrivée à Auckland!

Hier soir, nous avons vu un bon film en anglais au cinéma. C'était marrant mais Rosie ne l'a pas aimé, qui est malheureux. Le point le plus marquant était quand Rosie est tombée avec son popcorn, car on a rigolé pour cinq minutes! On a hâte d'aller à la Skytower et avoir bien ensemble demain.

1. I wrote half a page (about activities I did with my exchange student) with no resources and then half a week later I went and checked it and corrected my mistakes in preparation for this portfolio writing. I should have done more than just this, but I was busy with a lot of other things the week before the test. It helped give me ideas for writing in the test, but I didn't bother memorising it because I think that's a bit pointless. I would like to be able to write in French well on the spot as opposed to being able to vomit a paragraph that I memorised the night before. Next time, I will study some common phrases that are a little tricky to say, like "we had lots of fun together" etc.
2. I didn't know some of the things I wanted to say, like the word(s) for "everything", so I had to make educated guesses. For "everything" I wrote "tous les choses" which I think is actually supposed to be "toutes les choses" but I didn't expect it to be correct at the start. However, minor mistakes in the first draft don't concern me at all, and most of my mistakes were minor and I will be able to correct them using the resources suggested. I think my story was well written, excluding the grammatical errors.
3. V - 2. L - 3. A. I am not surprised and I find the codes really helpful. I think this way of marking is a really good idea because I always learn lots with the codes. Teaching on common errors was great too, but it's quite quick and I forgot most of the tips already. However, it's briefly summarised on a sheet so I can look back at that. I prefer the codes but both were helpful.
4. The comment isn't helpful but the codes indicate what needs to be fixed and I am happy with that.

Appendix M: Participant Information Sheet and Assent Form

Faculty of Education
School of Curriculum and Pedagogy
Epsom Campus
+64 9 623 8899
The University of Auckland
Private Bag 92019
Auckland, New Zealand

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

PARTICIPANT INFORMATION SHEET
Student

Project title: First Language Reflection for Second Language Development
Name of Researcher: Anne Moir Scott

Researcher Introduction
As well as being a French teacher, I am a Doctor of Education student in the School of Curriculum and Pedagogy in the Faculty of Education at the University of Auckland. This study is being conducted for my doctoral thesis. The Principal of the school has given permission for this research to take place and has given assurance that your participation or non-participation in the research will not affect your grades.

Project description and invitation
The study investigates the effectiveness of reflection when adolescents write in French Foreign Language for NCEA portfolio assessment. Data gathering will take place in Term 1 and Term 2, 2015 and the study will last until Dec 2015.

I am approaching you, as a Year 11 French student, for your permission to use your written French and accompanying reflections as data for analysis. Your writing and reflections will be your usual classwork and no extra time will be required of you than is normally required in the Year 11 programme. Apart from giving your consent to supply your written French and reflections, nothing further will be required of you.

Project Procedures
I will explain the project to you and be available to answer any of your questions. If you agree to take part in this study, please sign the attached assent form. Please be assured that participation in this study is entirely voluntary (you do not have to assent to your coursework/reflections being accessed for research purposes). The data for the research will come from your first and second drafts of two tasks (four drafts per student) written under test conditions in two process writing units. In one of the two writing units, you will be given the opportunity to reflect in an online journal on the processes and strategies you use as you write in French. Although your name, texts and personal data will be known to me they will kept confidential to me as the researcher.

Data storage/retention/destruction/future use
The French drafts will be stored in the school in the usual manner for NCEA writing and any research data relating to them will be stored securely. I will set up an online blog site and give access to the students to enter their reflections. These reflections will be visible to the individual students, their teachers and to me only. The reflective data will be downloaded into a text document for analysis and access to the blog will be terminated at the end of the school year. The consent forms will be stored separately in a lockable cabinet in the office of my main supervisor, Associate Professor Martin East. All data will be stored for six years from the commencement of the project, then destroyed. All paper data will be shredded and electronic data deleted from the computer on which they are stored.

Optional Participation
Participation in this research is entirely voluntary for you and after giving assent, you may withdraw at any time.

Anonymity and Confidentiality
Data from the project will be used for my Doctor of Education thesis and may result in academic presentations or publications. However, no identifying information will be included in any reporting of the findings of the study and the students' anonymity in reporting is therefore assured. Pseudonyms will be used in any reporting/publications. A summary of the research will be made available to you. Please indicate in the Assent Form if you would like to receive a copy.

Faculty of Education
School of Curriculum and Pedagogy
Epsom Campus
+64 9 623 8899
The University of Auckland
Private Bag 92019
Auckland, New Zealand

If you have any queries or wish to know more please ask me in the first instance. You can also contact my supervisor and the Head of School. Contact details are:

Name: Anne Moir Scott
Email: amoi003@aucklanduni.ac.nz
Phone: 09 9706726

My supervisor:

Associate Professor Martin East,
School of Curriculum and Pedagogy, Epsom Campus,
Faculty of Education,
The University of Auckland,
Private Bag 92019,
Auckland
Phone: +64 9 623 8899 Ext 48692.
He can be contacted by email: m.east@auckland.ac.nz

The Head of School:

Professor Judy Parr,
School of Curriculum and Pedagogy,
Epsom Campus,
Faculty of Education,
University of Auckland,
Private Bag 92019, Auckland,
Phone: +64 9 623 8899 Ext 88998.
Her email: jm.parr@auckland.ac.nz.

For any queries regarding ethical concerns you may contact:

The Chair,
The University of Auckland Human Participants Ethics Committee,
The University of Auckland,
Research Office,
Private Bag 92019, Auckland 1142.
Telephone 09 373-7599 extn 87830/83761
Email: humanethics@auckland.ac.nz

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 19TH DECEMBER 2014 FOR THREE YEARS, REFERENCE NUMBER 013442

Faculty of Education
School of Curriculum and Pedagogy
Epsom Campus
+64 9 623 8899
The University of Auckland
Private Bag 92019
Auckland, New Zealand

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

ASSENT FORM
Student

THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS

Project title: First Language Reflection for Second Language Development
Name of Researcher: Anne Moir Scott

I have read the Participant Information Sheet, have understood the nature of the research and why my school and I have been selected. I understand that my Principal has given permission for this research to take place and has given assurance that participation or non-participation in the research will not affect my grades. I have had the opportunity to ask questions and have them answered to my satisfaction.

☐ I volunteer to take part in this research in terms of granting access to my regular coursework / reflections on writing.

☐ I understand that participation or non-participation in the research will not affect my learning and assessment (grades awarded).

☐ I understand that I am free to choose not to participate or to withdraw participation at any time after giving consent.

☐ I understand that the texts/reflections will not be anonymous but that identities will be kept confidential to the researcher. Furthermore, although data/findings may be used in presentations or publications, no identifiable information about me or any other student will be made available.

☐ I understand that data will be kept for six years, after which they will be destroyed.

☐ I wish / do not wish to receive a summary of the research findings.

Name: _____

Signature: _____ Date: _____

Contact detail if you wish a copy of the final research report: _____