A meta-investigation of the use of the labels 'longitudinal' and 'long-term' in studies of feedback on writing

Dr. Rachael Ruegg, *Victoria University of Wellington, New Zealand*
Dr. Paul Iida, *Akita International University, Japan*

Abstract

A number of studies have called for more 'longitudinal' research of feedback on writing. However, few offer concrete definitions of the term and in practice it seems to be used inconsistently. The purpose of this meta-investigation was to explore how the terms 'longitudinal' and 'long-term' are used within the literature on feedback on writing in order to determine what the terms mean in this context. Understanding the ways in which the terms are used will increase clarity in terms of the extent of longitudinal research in the area of feedback on writing and the extent to which further longitudinal research is still needed. Having a clearer understanding of the ways in which the terms are used will help researchers to design research to fill the reported gap. Analysis illustrated discrepancies in the use of the words: a wide range in the length of time, vast difference in the amount of feedback and the number of times feedback was given. It may be prudent for researchers to not only describe their research design using labels, but to also justify the basis on which each label applies to the research.

Keywords: *Longitudinal; long-term; feedback on writing; research methodology; research design*

Introduction

Researchers are requested to describe their research design in great detail when publishing empirical studies in journals in the field of applied linguistics. One aspect of this description is the timeframe over which the study was conducted. Many textbooks on the subject of research design juxtapose short-term 'cross-sectional' research to long-term 'longitudinal' research (e.g. Dörnyei, 2007; Phakiti, 2014). Cross-sectional research represents a snapshot of the situation by comparing groups of students (for example, at different proficiency levels) at a single point in time, whereas longitudinal research represents "an ongoing investigation of people or phenomena over time" (Dörnyei, 2007, p. 78). A great number of studies have called for more longitudinal research in the field of applied linguistics in general, not least in the study of feedback on writing (e.g. F. Hyland, 2010; Van Beuningen, 2010; Storch, 2010). Second Language Acquisition (SLA) theory suggests that interlanguage development happens gradually over a long period of time (e.g. Gass, 2003). Therefore, measurable improvement after a single feedback treatment seems unlikely.

Whereas, longitudinal data can provide greater evidence for an effect on writing, rather than relying on a single iteration of feedback (Wildemuth, 2016). However, few research methods materials offer concrete definitions of the term and in practice it seems to be used inconsistently to describe many different kinds of research designs.

The purpose of the present study was to investigate the ways in which the terms 'longitudinal' and 'long-term' are used within the literature on feedback on writing in order to determine what the terms mean in this context. Understanding the ways in which the terms are used will provide increased clarity in terms of the extent of longitudinal research in the area of feedback on writing and the extent to which further longitudinal research is still needed. Having a clearer understanding of the ways in which the terms are used will help researchers to design research to fill the reported gap. The research question for the study was: What kinds of research designs are described by the use of the terms 'longitudinal' and 'long-term' in the area of feedback on writing?

Defining Longitudinal Research

The fundamental concept used in definitions of the word 'longitudinal' is the collection of the same types of data, from the same learners over a period of time. However, a period of time is often not defined in concrete terms in research methodology literature and may vary depending on what type of data is being collected and the purpose of the analysis (Menard, 2002, p. 2). For research in the field of applied linguistics, most writers tend to agree that the minimal period of time which can be described as longitudinal is a few months (Guenette, 2007; Mackey & Gass, 2012; Phatiki, 2014). Although, as explained by Menard (2002), cross-sectional research investigates "few periods, regardless of the actual length of a single period" while longitudinal research investigates many periods (p. 50).

Longitudinal research involves the collection of the same data from the same participants, or participants that are comparable in every way (Guenette, 2007), over multiple time periods. The same data must then be matched for the purposes of analysis (Phatiki, 2014). The data are compared to measure change in a particular variable/s, such as interlanguage development or change in the relationship between two or more variables (Menard, 2002). In addition to the traditional 'panel study,' a cross-sectional design can be conducted repeatedly in order to discover trends, and longitudinal research can also be conducted retrospectively by collecting data at a single period of time and drawing on participants' memory of past events (Dörnyei, 2007).

Not all writers, even within the field of applied linguistics, agree on the boundaries of what constitutes longitudinal research. Dörnyei (2007) claims that "both ethnography and case study research emphasize prolonged engagement with the participants…, making them inherently longitudinal" (p. 81), while Phatiki (2014) states that "it is necessary to distinguish longitudinal research from prolonged research and extensive data triangulation techniques over time – the typical research characteristics of case studies and ethnographies" (p. 10). If even research design experts cannot agree on what constitutes longitudinal research, how are researchers to determine whether their research should be considered to be longitudinal or not?

Advantages of Longitudinal Research

A large amount of longitudinal research has been conducted across a wide range of academic fields. Despite this method being a relatively common approach to answering research questions and testing hypotheses, little criticism has ever been levelled at longitudinal research design per se (Menard, 2002, p. 1).

Collecting the same data from the same individuals at different points in time enables us to not only quantitatively measure changes, but also to qualitatively describe the trajectory of such changes. Duff (2006) argues that longitudinal studies can reveal various developmental pathways taken by different learners. Some scholars even suggest that longitudinal studies "provide the strongest evidence in support of developmental patterns" (Ellis, 2003, p. 75), as well as patterns of error production that continue over a longer time period (p. 55). Therefore, longitudinal research is useful for investigating changes in language learning or behaviour (Phatiki, 2014) as well as being the ideal method to track changes in a learner's interlanguage (Mackey & Gass, 2012) development over time.

Clearly, development is of utmost importance within the field of education in general and the possibility of measuring language development makes longitudinal research one of the most promising research directions for researchers in the field of applied linguistics (Abbuhl & Mackey, 2008). Especially, studies which are designed to measure the effects of pedagogical practices are unlikely to find substantive development using a one-shot research design.

Lack of Longitudinal Studies of Feedback on Writing

Despite the stated benefits of longitudinal research, in the field of second language acquisition, Ellis (2003) points out that there are a limited number of longitudinal studies

covering a variety of areas of study within SLA, such as error production (e.g., Chamot, 1978, 1979), language development (e.g., Schmidt, 1983), morphemes (e.g., Dulay & Burt, 1974; Hakuta, 1974; Rosansky, 1976), the setting where language learning takes place (e.g., Klein & Dittmar, 1979; Meisel, 1983; Schmidt, 1983; Schumann, 1978), and even the sex of the participants (e.g., Burstall, 1975). Ellis' findings are that the majority of studies are cross-sectional, with a single point of data collection. More specifically, many writers on the subject have commented on a distinct lack of longitudinal research on issues relating to feedback on writing. Liu and Hansen (2002) mentioned a lack of research on the long-term benefits of peer feedback, while others (Ferris, 2003; 2010; Guenette, 2007; Hyland & Hyland, 2006; F. Hyland, 2010) have noted a lack of longitudinal research on issues related to feedback on writing more generally. Furthermore, Dörnyei (2007) laments the lack of longitudinal studies in research conducted in the field of applied linguistics as a whole and "…present[s] a strong case for the need to experiment more with longitudinal designs in investigations" (p. 78).

Guenette (2007) specifically calls for more longitudinal studies measuring "…the development of accuracy over time" (p. 44). On the other hand, Hyland and Hyland (2006) suggest "…research designed to understand the longitudinal effects of teacher comments on student writing, focusing on questions such as what types of feedback lead to writing development over time and whether revisions to drafts [lead to] improvement in later writing…" (p. 96). F. Hyland (2010) argues for qualitative longitudinal research conducted in naturalistic settings that focus on the engagement of individual learners with the feedback they receive over a complete course. Although a complete course could represent a wide range of time periods, it would most often indicate one semester or one academic year.

Previous Studies Identified as Longitudinal by Others

Several writers who lamented a lack of longitudinal research in the L2 writing field also identified studies which they considered to be longitudinal in nature, despite the fact that the articles reporting the studies in some cases had not identified them as longitudinal. Kepner (1991) uses the word 'longitudinal' to describe her study and her study is also identified as longitudinal by Ferris (2003). Guenette (2007) mentions six studies that have "…traced the development of accuracy over time" (p. 44; cf. Chandler, 2003; Fazio, 2001; Kepner, 1991; Polio, Fleck, & Leder, 1998; Semke, 1984; Sheppard, 1992). Of these, two describe their studies as 'long-term' (Chandler, 2003) or 'longitudinal' (Kepner, 1991), while the remaining four studies do not self-identify as such. Similarly, Bitchener, and Ferris (2012) point out six studies "of long-term effects of feedback" (p. 88; cf. Chandler, 2003; Ferris, 1995; 2006;

Ferris, Liu, Sinha, & Senna, 2013; Foin & Lange, 2007; Haswell, 1983), one of which describes itself as 'longitudinal' (Ferris, Liu, Sinha, & Senna, 2010), one of which self-identifies as 'long-term' (Chandler, 2003), and one of which uses both terms to describe its methodology (Ferris, 2006), while the remaining three articles use neither 'longitudinal' nor 'long-term' to describe the research conducted (Haswell, 1983; Ferris, 1995; Foin & Lange, 2007).

As seen above, there are some inconsistencies and certain ambiguities in the definition of longitudinal research in the research methodology literature. The specific focus of feedback on writing was chosen for the investigation because, as noted in this section, there has been reported to be a distinct lack of longitudinal or long-term research in this particular area of second language education. It was also decided that choosing a narrow area within the larger field of language learning would afford the researchers the ability to provide a clearer analysis. There are insufficient longitudinal or long-term research studies on feedback on writing employing quantitative research methods. Therefore, a quantitative meta-analysis was not possible. Consequently, this study employs qualitative meta-investigation to understand the ways in which longitudinal and long-term research are operationalized within the specific area of feedback on writing, in order to make the meaning of the words more tangible. It is hoped that this study will raise awareness of the lack of longitudinal studies within this area of applied linguistics and contribute to a clearer understanding of the kinds of research that would fill this gap.

Materials and Methods

Literature Search

In order to locate all of the relevant studies related to the research question, the researchers conducted a preliminary search of online databases (JSTOR, ERIC, and Proquest), as well as the researchers' own personal collections of books and journal articles and secondary sources which were referred to in other sources, using combinations of the following search terms: 'longitudinal' and 'long-term,' coupled with the terms 'writing' and 'feedback.' The initial results generated 52 studies for consideration, which were later filtered based on the criteria for inclusion described in the following section.

Criteria for Inclusion

The first criterion for inclusion in the present study was that each source had to either explicitly use one of the terms 'longitudinal' or 'long-term' in the description of its procedures or methods, or had to be identified by a secondary source as a longitudinal or long-term

study. For example, in their book on Written Corrective Feedback (WCF), Bitchener and Ferris (2012) identify six previous studies measuring long-term effects of WCF.

Additionally, each study had to have as its focus an examination of the impact that feedback of any description has on writing in the context of a language or composition classroom. The researchers thoroughly read each of the 52 studies to determine whether they met these criteria. Once the search was complete, the researchers met to discuss any discrepancies in their findings and make the final decision about which studies to include and exclude based on the criteria. In 28 of the 52 cases, the studies did not meet all of the criteria and were excluded from the analysis. Because only 24 studies were found that met the study criteria, it was decided not to narrow the data further based on the context of the research reported. Thus, studies conducted at all educational levels and in all languages were included in the analysis. The final sample included 18 studies focusing on writing in English, and 6 focusing on other languages. It included 19 studies carried out in university contexts, and 5 either carried out in other contexts, or that failed to report this information.

Coding and Analysis

In order to answer the research question, the 24 remaining studies were coded independently by each of the two researchers using the study characteristics described in the next paragraph. When the independent coding was complete, the researchers met to compare notes and look for discrepancies in their coding. Once the discrepancies were addressed, the coding process was complete.

The following characteristics were identified as key factors to analyse how researchers view the concept of longitudinal research design: which term was used to describe the study (i.e., longitudinal or long-term), the description of their term choice (e.g., 16-week semester, 10 months), the actual length of time of the study from start to finish, the period of time that feedback was given, the number of feedback iterations given, the amount of feedback, the methodological orientation of the study (i.e., quantitative or qualitative), and whether the feedback was provided orally, in written form, or both. At this stage, one further study (Matsumura, Patthey-Chaves, Valdes, & Garnier, 2002) was excluded from the analysis, although it met the initial criteria for inclusion, because of a lack of details reported in the article about the research methods used. Indeed, the article only included information about four of the eight characteristics investigated.

Upon completion of the coding, the researchers approached the data through the lens of content analysis (CA). This study adopted Neuendorf's (2011) definition of CA as a 'systematic, objective, quantitative analysis of message characteristics' (p. 1). Krippendorff (2013) adds that CA is one of the most significant methods for analysing data in the social sciences, because such analysis affords the researcher the ability to view text from a unique perspective that other methods do not offer (p. xii). He explains that CA: 1) is 'an empirically grounded method' (p. 1) that methodologically studies textual data, how individuals interpret that data, as well as the impact that such data has on society; 2) 'transcends traditional notions of symbols, contents and intents' (p. 2) where texts are more than mere words, but that they contain deeper meanings and the words are simply a vehicle through which these meanings are transferred from one individual to another; and 3) gives researchers the means by which to critically examine data, no matter the outcome, through a unique analysis (pp. 1-5).

In the context of this study, CA afforded the researchers the ability to examine not only the surface-level choice that researchers made in describing their research design, but the implications of that choice. In particular, the researchers were interested in understanding the underlying message that is conveyed when studies are described as longitudinal or long term. In an effort to comprehend these messages, the researchers reviewed the coded data following the approach of constantly comparing data throughout the analysis process (Glaser, 1992; Glaser & Strauss, 1967). This comparison provided the researchers with the opportunity to look for themes to emerge from commonalities in the data. The initial themes that surfaced were: term used, explanation/definition of the term, length of the research from beginning to end, feedback period, number of feedback iterations, amount of feedback given, research orientation and mode of feedback. As analysis continued, themes were collapsed, divided or otherwise revised until the data analysis no longer required changes. This method resulted in the following final themes: Term used and description of the term, length of research from beginning to end, period of time and frequency of feedback, amount of feedback, research orientation (quantitative or qualitative) and form of feedback (oral, written, or a combination).

Results

Table 1 shows all of the studies included in the meta-investigation and identifies the ways in which they were classified for the purposes of this research. The studies that did not self-identify as 'longitudinal' or 'long-term', but which were identified as longitudinal or long-term studies in other books or journal articles, have no information included in the second and

third columns. The word ‘interview’ is used to explain a conference about writing held between a researcher and a student, whereas the word ‘conference’ is used to explain a discussion of writing between a teacher and student. The results are discussed in more detail below.

Table 1. *Coding of studies included in this study.*

Study	Term used	Description	Length of research	Period of feedback	Feedback iterations	Amount of feedback	Form of feedback	Research orientation
Ahmadi-Azad (2014)	longitudinal (effects)	-	9 weeks	4 weeks	4	NR	direct/indirect written feedback + oral feedback	Quantitative
Benevento & Storch (2011)	longitudinal	6 months	6 months	6 months	3	comprehensive (grammar + vocabulary)	WCF	Quantitative
Bitchener & Knoch (2009)	longitudinal	10 months	10 months	1 day	1	NR	Mostly WCF	Quantitative
Chandler (2003)	long term (error)	Semester	Semester	Semester	10	comprehensive (grammar + vocabulary)	written	Quantitative
Degteva (2011)	longitudinal	7 weeks	7 weeks	7 weeks	9	NR	WCF	Quantitative
Fazio (2001)	-	-	5 months	3.5 months	Around 14	NR	written	Quantitative
Ferris (1997)	-	-	one semester	¾ semester	6	13.87 points (mean)	comments	Descriptive statistics
Ferris (2006)	longitudinal (design), long-term (effects)	semester	15 weeks	NR	4	61.66 points (mean)	WCF	Quantitative
Ferris, Liu, Sinha & Senna (2013)	longitudinal	semester	16 weeks	12 weeks	6	51.5 points (mean)	WCF followed by TSC	Qualitative

Geilen, Tops, Dochy, Onghena, & Smeets, (2010)	long-term	2 trimesters	6 months	4 months	3	NR	peer & teacher written feedback	Quantitative
Haswell (1983)	-	-	one semester	one semester	NR	16.7 points (mean)	indirect written feedback	Quantitative
Hyland (1998)	longitudinal	14 weeks	14 weeks	14 weeks	12	109 points (mean)	peer & teacher written and oral	Descriptive statistics
Kepner (1991)	longitudinal	Semester	12 weeks	NR	6	NR	written	Quantitative
Lee & Schallert (2008)	long	"sufficiently long"	6 weeks	NR	NR	NR	written	Qualitative
Nordrum, Evans & Gustaffson (2013)	Long term (learning)	NR	14 weeks	7 weeks	5	NR	written	Qualitative
Polio, Fleck & Leder (1998)	-	-	15 weeks	7 weeks	7	NR	written	Quantitative
Poverjuc, Brooks & Wray (2012)	longitudinal	NR	1 year	NR	6-9	35-87 minute interview	oral interview	Qualitative
Riazantseva (2012)	longitudinal	semester	14 weeks	NR	weekly	NR	written	Quantitative
Robb, Ross & Shortreed (1986)	-	-	9 months	NR	NR	NR	written	Quantitative
Semke (1984)	-	-	10 weeks	9 weeks	9	NR	written	Quantitative
Seror (2011)	longitudinal	eight-month	8 months	8 months	around 16	1 hour interview	oral interview	Qualitative
Sheppard (1992)	-	-	10 weeks	7 weeks	7	NR	written	Quantitative

Simpson (2006)	long-term (improvements)	semester	16 weeks	NR	around 10	NR	WCF and/or comments	Qualitative
Vyatkina (2010)	long-term (error-rate changes)	one semester	16 weeks	NR	NR	7.36 points (mean)	written	Quantitative

Term Used and Description of the Term

Out of the 24 studies included in this analysis, 17 self-identified as longitudinal or long-term studies. Ten used the term 'longitudinal', and one of those more specifically focused on the 'longitudinal effects of feedback'. A further five studies used the term 'long-term'; more specifically, one focused on 'long-term improvements', one looked at 'long-term learning', one measured 'long-term error rate changes', and one considered 'long-term error'. In addition to this, included in this analysis was one study that used both 'longitudinal' and 'long-term,' mentioning that the research employed a 'longitudinal design' and focused on the 'long-term effects of feedback'. There was also one study that used the term 'long,' mentioning that the duration of the study was 'sufficiently long to develop a deep understanding of the development of the relationships between teacher and students in context' (Lee & Schallert, 2008, p. 516).

Apart from these 17 studies which used one of the terms to refer to their own study, there were also seven studies which were referred to by another source as either focusing on the 'long-term effects' (Haswell, 1983, as cited in Bitchener & Ferris, 2012) or 'long-term gains' (Robb, Ross, & Shortreed, 1986, as cited in Ferris, 2006), or constituting 'longitudinal' research on the topic (Fazio, 2012, as cited by Guenette, 2007; Ferris, 1997, as cited in Matsumura, Pathey-Chaves, Valdes, & Garnier, 1997; Polio, Fleck, & Leder, 1998, as cited by Guenette, 2007; Semke, 1984, as cited by Guenette, 2007; Shepphard, 1992, as cited by Guenette, 2007). Figure 1 illustrates the break-down of the difference in terminology.

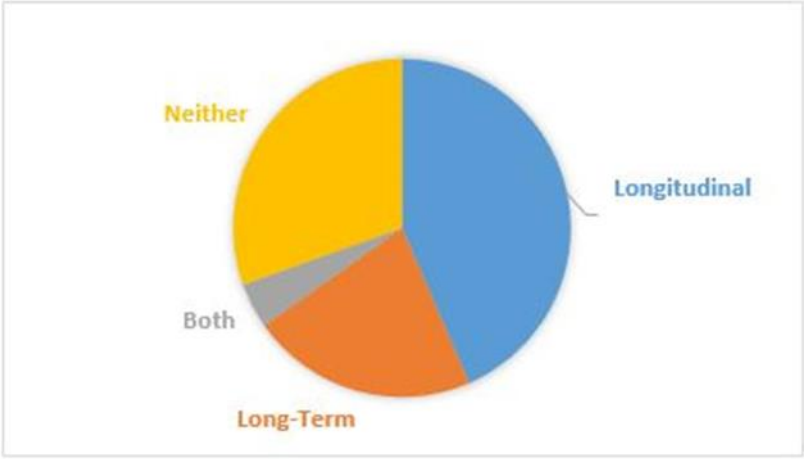

Fig. 1. *Breakdown of term usage*

Length of the Research from Beginning to End

Three of the 24 studies analysed were conducted over a period of less than one semester, despite referring to themselves as either 'longitudinal' or 'long.' The shortest study was

conducted over a period of six weeks. One study was conducted over a period of seven weeks and another one was conducted over nine weeks. The majority of the studies (15 out of the 24 analysed) were conducted over a period of one semester, although the length of the semesters ranged from ten weeks to 16 weeks.

Five of the studies continued for more than one semester, but less than one year. Of those, two studies were conducted over a 6-month period, while another three were conducted over eight to ten months. Finally, the longest of the 24 studies analysed was conducted over a period of one year. Figure 2 illustrates the break-down of length of time for the studies.

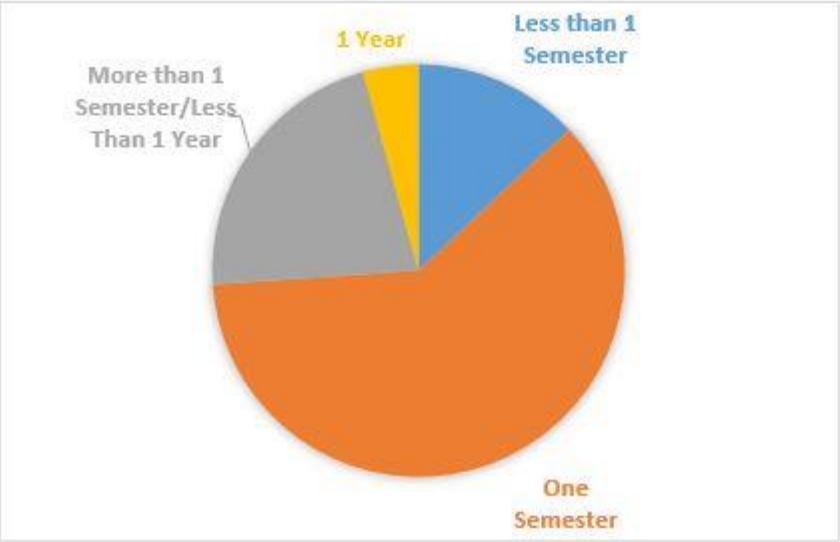

Fig. 2. *Break-down of length of time*

Period of Time and Frequency of Feedback

In terms of the period of time over which feedback was provided to students, a large number of studies (8 out of 24 studies) did not provide any information. Out of the 16 studies that did specify the feedback period, nine provided feedback over a period of less than one semester. This included one which provided feedback on only one day, one which provided feedback over a period of four weeks, four which provided feedback over a period of seven weeks, one which provided feedback over a period of nine weeks, one over a period of 11 weeks and one over a period of 12 weeks. In addition to this, there were five studies that involved the provision of feedback over a period of one semester, or more specifically, usually 14 to 16 weeks. Finally, there were two studies in which feedback was provided for a period of more than one semester. It was provided over a period of six months in one and a period of eight months in the other. See Figure 3 for a description of the length of time that feedback was provided.

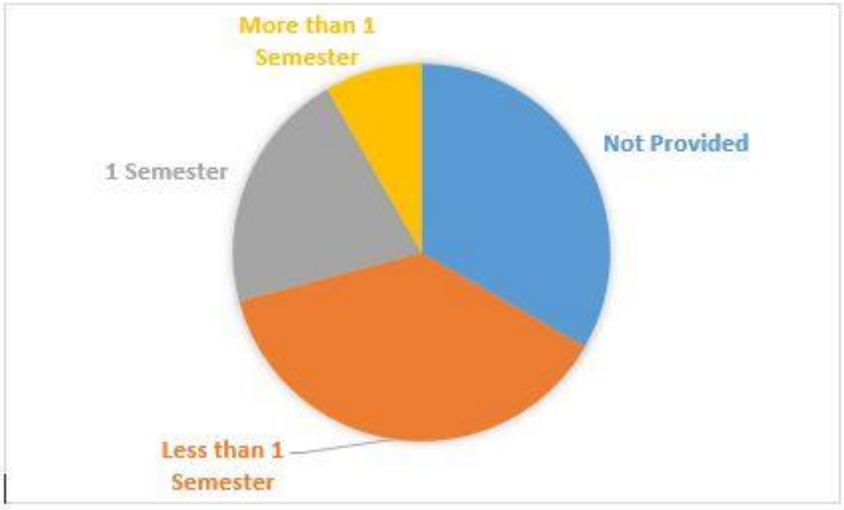

Fig. 3. *Length of time for the provision of feedback*

There were four studies that did not specify how many times feedback was provided. Out of the 20 studies that did provide information on the number of feedback iterations, six studies provided feedback one to five times. Included in this number are one study which involved the provision of feedback only once, two studies which involved the provision of feedback three times, two which involved the provision of feedback four times, and one which entailed the provision of feedback five times. Additionally, there were ten studies that involved six to ten feedback iterations. Finally, the data pool included four studies in which feedback was provided more than ten times: one in which it was provided 12 times, two in which it was provided 14 times, and one in which it was provided 16 times. See Figure 4 for an illustration of the number times feedback was provided during each study.

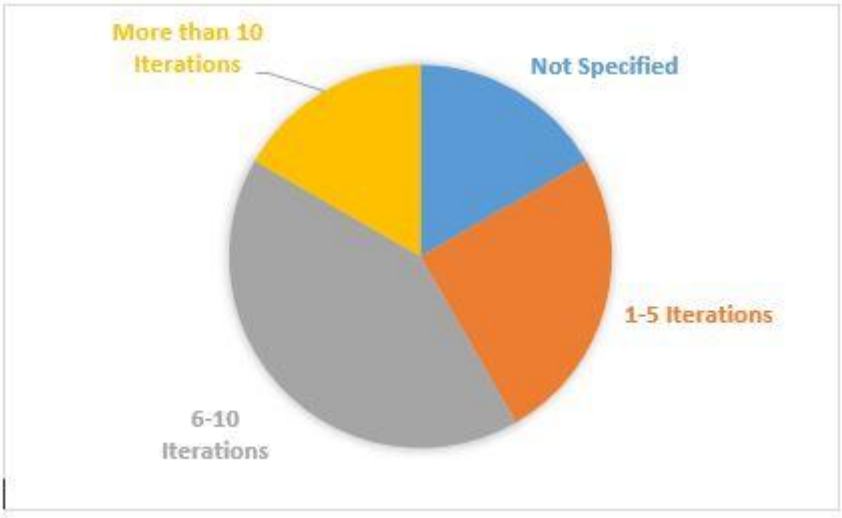

Fig. 4. Number of iterations

Although the range is more varied in the 'longitudinal' studies, difference in the average number of iterations is negligible (an average of about 8 iterations for the 'longitudinal' studies and about 7 for the 'long-term' studies).

Amount of Feedback

The amount of feedback provided in each study was more difficult to compare, since each study described the amount of feedback in a different way. Two studies involved the provision of comprehensive feedback, but no examples were provided so it was difficult to imagine how much was actually given. Furthermore, one study involved teacher student interviews, with each student taking part in a one-hour interview with the teacher. Some studies provided descriptive statistics that quantified the amount of feedback provided, while two provided examples in the appendices that the researchers in this study quantified. Moreover, 14 of the 24 studies analysed provided no information whatsoever about the amount of feedback provided on each draft.

Of the ten studies that did provide information, three studies provided comparatively less feedback in each iteration. Vyatkina (2010) provided around 7-8.5 instances of feedback per 100 words, Ferris (1997) provided 13.87 instances of feedback on each student's draft and Haswell (1983) provided on average 16.7 instances of feedback on each draft. The remaining seven studies provided comparatively more feedback. Seror (2011) included a one hour interview with each student. Two studies provided comprehensive feedback (Benevento & Storch, 2011; Chandler, 2003). Poverjuc, Brooks, and Wray (2012) conducted interviews which lasted from 35 to 87 minutes with each student. Ferris, Liu, Sinha, and Senna (2013) provided on average 51.5 instances of feedback on each draft, Ferris (2006) provided an average of 61.66 instances of feedback per draft, and Hyland (1998) provided over 100 instances of feedback on each draft on average (a range of 44 to 206 individual instances of feedback per draft).

Research Orientation and Form of Feedback

Eight of the 24 studies took a qualitative perspective. The remaining 16 studies took a quantitative perspective. A vast majority of the studies (19 out of the 24 analysed) involved providing only written feedback to the students. Two studies involved the provision of oral feedback alone. The remaining three studies involved a combination of oral and written feedback being provided to the students.

Discussion

It may be surprising to many instructors and researchers to see that, even amongst research that is defined as 'longitudinal' or 'long-term' only one of the 24 studies analysed lasted for a period of a year. Furthermore, studies which last for less than one semester would usually not meet F. Hyland's (2010) suggestion of investigating feedback over an entire course, as most courses last for at least one semester. It is also interesting to note that no studies lasting less than a semester were referred to as 'longitudinal' or 'long-term' by others. On the other hand, four studies lasting for less than one semester were self-described as either 'longitudinal' (n = 3) or 'long-term' (n = 1).

Menard (2002), explains that cross-sectional research investigates few periods, irrespective of the time involved while longitudinal research investigates many periods (p. 50). The number of feedback iterations may go some way to providing an insight into the number of individual instructional or learning periods included in longitudinal research on feedback on writing. Of the studies analysed in this meta-investigation, five involved the provision of feedback between one and four times, which may be considered by most to be few. On the other hand, seven studies involved providing feedback ten times or more, which would not be considered few by most instructors and researchers of writing. The largest number of studies (ten of the 22 that included the information) fell somewhere in between these two extremes.

Although there are no clear concrete definitions of the word 'longitudinal' in the research methodology literature, there are some apparent discrepancies between the rough description of 'longitudinal' research in the research methodology literature and the use of the word in studies of feedback on writing. Analysis of the studies included in this research have illustrated these discrepancies, with the wide range in the length of time from start to finish, from just a few weeks to more than one year. The vast difference in the amount of feedback provided and the number of iterations may raise questions as to the quality and purpose of designing longitudinal studies that examine the effects of feedback on writing.

A problem that arose in the analysis of the studies included in this research was that several studies failed to include some of the basic information that is typically expected when describing the methodology of a research study. For example, in reviewing the details of each of the 24 studies examined, one third of them did not report the length of time during which feedback was given to students. Another problem was that several studies failed to describe the number of iterations provided to students, which is also an important

component for understanding the effectiveness of feedback on learners' writing. Finally, many of the studies (ten of the 24 examined) did not reveal how much feedback they provided to students, which again, is a significant factor in determining the effectiveness of feedback. Not only was the lack of reporting a problem, but of those that did report the amount of feedback, how this information was reported varied significantly from study to study, illustrating the lack of uniformity in describing the methods employed in studies.

There are many aspects of research methodology that researchers are required to justify in the publication of their research results. However, a researcher is rarely required to justify the use of a label, such as 'longitudinal' or 'long-term'. In terms of improving the quality of research publications in the field of applied linguistics in general, there is a continued need for researchers who publish their results in academic journals to go to great lengths to describe, explain, illustrate, and justify all aspects of their research design so that readers have all the information necessary to understand the implications of the results. It may also be prudent for editors and reviewers of journal articles in the field to ask researchers to not only describe their research design using the labels, but to also justify the basis on which each label applies to the research being reported. Over the long term, this increased detail in research reports will provide a rich illustration of such labels and what they really mean when applied to experiments and naturalistic research in our field.

Conclusion

The main purpose of this study was to understand what the terms 'longitudinal' and 'long-term' mean in the context of studies of feedback on writing. Based on the results of this study, it would seem that both terms are used to describe research that is conducted over a period of at least one semester, as only four of the 24 studies were conducted over shorter periods. This corroborates F. Hyland's (2010) suggestion that longitudinal research should be conducted over the period of an entire course, as most courses are one semester in length. The results of this study also suggest that research of feedback on writing should be considered longitudinal if it consists of four feedback iterations or more, as only three of the 24 studies involved less than four feedback iterations. Furthermore, cross-sectional research investigates few periods, while longitudinal research investigates many periods (Menard, 2002, p. 50). Three feedback iterations would be considered as 'few' by most of the studies included in this analysis. Therefore, this research corroborates Menard's (2002) description.

In order to improve the quality of reporting methods used in conducting research, the findings of this study support previous calls for increased standardisation on what is reported

and how it is reported (DeKeyser & Schoonen, 2007; Norris, Plonsky, Ross, & Schoonen, 2015). Not only will greater uniformity make the reporting of research more reader-friendly, but such consistency in reporting findings on writing feedback, as well as other areas of language learning, will help researchers to replicate studies with greater accuracy and to fully understand how various factors may influence results.

Biodata

Dr. Rachael Ruegg is a senior lecturer in the School of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand. Her research interests include feedback on writing, assessment of writing and writing support.

Dr. Paul Iida is a professor of English for Academic Purposes at Akita International University, Japan. His research interests include instructed SLA and issues of marginalized learners and teachers in the K-16 environment.

References

Abbuhl, R., & Mackey, A. (2008). Second language research methods. In K. King & N. Hornberger (Eds.), *Encyclopedia of language and education* (Vol. 10, Research Methods, pp. 99-111). New York, NY: Springer.

Ahmadi-Azad, S. (2014). The effect of coded and uncoded written corrective feedback types on Iranian EFL learners' writing accuracy. *Theory and Practice in Language Studies, 4*(5), 1001-1008. doi:10.4304/tpls.4.5.1001-1008

Benevento, C., & Storch, N. (2011). Investigating writing development in secondary school learners of French. *Assessing Writing, 16*(2), 97-110. http://dx.doi.org/10.1016/j.asw.2011.02.001

Bitchener, J., & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing.* New York, NY: Routledge.

Bitchener, J., & Knoch, U. (2009). The value of a focused approach to written corrective feedback. *ELT Journal, 63*(3), 204-211. https://doi.org/10.1093/elt/ccn043

Burstall, C. (1975). Factors affecting foreign-language learning: A consideration of some relevant research findings. *Language Teaching and Linguistics Abstracts, 8*, 105-125. https://doi.org/10.1017/S0261444800002585

Chamot, A. (1978). Grammatical problems in learning English as a third language. In E. Hatch (Ed.), *Second language acquisition* (pp. 175-189). Rowley, MA: Newbury House.

Chamot, A. (1979). Strategies in the acquisition of English structures by a child bilingual in Spanish and French. In R. Andersen (Ed.), *The acquisition and use of Spanish and English as first and second languages* (pp. 90-106). Washington, DC: TESOL.

Chandler, J. (2003). The efficacy of various different kinds of feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*, 267-296. http://dx.doi.org/10.1016/S1060-3743(03)00038-9

Degteva, O. (2011). Impacts of recasts on the accuracy of EFL learners' writing (Unpublished Master's thesis). Kyrenia, Northern Cyprus: Girne American University.

DeKeyser, R., & Schoonen, R. (2007). Editors' announcement. *Language Learning, 57*, ix-x. doi:10.1111/j.1467-9922.2007.00396_2.x

Dörnyei, Z. (2007). *Research methods in applied linguistics.* Oxford, United Kingdom: Oxford University Press.

Duff, P. (2006). Beyond generalizability: Contextualization, complexity, and credibility in applied linguistics research. In M. Chalhoub-Deville, C. Chapelle, & P. Duff (Eds.), *Inference and generalizability in applied linguistics: Multiple perspectives* (pp. 65-95). Amsterdam, Netherlands: John Benjamins.

Dulay, H., & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning, 24*, 37-53. doi:10.1111/j.1467-1770.1974.tb00234.x

Ellis, R. (2003). *The study of second language acquisition*. Oxford, United Kingdom: Oxford University Press. (Original work published 1994)

Guenette, D. (2007). Is feedback pedagogically correct?: Research design issues in studies of feedback in writing. *Journal of Second Language Writing, 16,* 40-53. http://dx.doi.org/10.1016/j.jslw.2007.01.001

Fazio, L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing, 10,* 235-249. http://dx.doi.org/10.1016/S1060-3743(01)00042-X

Ferris, D. (1995). Can advanced ESL students become effective self-editors? *The CATESOL Journal, 41,* 41-62. Retrieved from http://www.catesoljournal.org/wp-content/uploads/2016/11/CJ8_ferris.pdf

Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315-339. doi:10.2307/3588049

Ferris, D. (2003). *Response to student writing: Implications for second language students.* Mahwah, NJ: Lawrence Erlbaum Associates.

Ferris, D. (2006). Does error feedback help student writers? In K. Hyland & F. Hyland (Eds.). *Feedback in Second Language Writing: Contexts and Issues* (pp. 81-104). Cambridge, United Kingdom: Cambridge University Press.

Ferris, D. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition, 32,* 181-201. https://doi.org/10.1017/S0272263109990490

Ferris, D., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing, 22,* 307-329. http://dx.doi.org/10.1016/j.jslw.2012.09.009

Foin, A., & Lange, E. (2007). Generation 1.5 writers' success in correcting errors marked on an out-of-class paper. *CATESOL Journal, 19,* 146-163. Retrieved from http://www.catesoljournal.org/wp-content/uploads/2014/07/CJ19_foin.pdf

Geilen, S., Tops, L., Dochy, F., Onghena, P., & Smeets, S. (2010). A comparative study of peer and teacher feedback and of various peer feedback forms in a secondary school writing curriculum. *British Educational Research Journal, 36(1),* 143-162. doi:10.1080/01411920902894070

Glaser, B. G. (1992). *Basics of grounded theory analysis: Emergence vs. forcing*. Mill Valley, CA: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York, NY: Aldine.

Guenette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing, 16,* 40-53. http://dx.doi.org/10.1016/j.jslw.2007.01.001

Hakuta, K. (1974). A preliminary report on the development of grammatical morphemes in a Japanese girl learning English as a second language. *Working Papers on Bilingualism, 3*, 18-43. Retrieved from http://files.eric.ed.gov/fulltext/ED123871.pdf

Haswell, R. (1983). Minimal marking. *College English, 45,* 600-604. doi:10.2307/377147

Hyland, F. (1998). The impact of teacher written feedback on individual learners. *Journal of Second Language Writing, 7*, 255-286. doi:10.1016/S1060-3743(98)90017-0

Hyland, F. (2010). Future directions in feedback on second language writing: Overview and research agenda. *International Journal of English Studies, 10*(2), 171-182. Retrieved from http://revistas.um.es/ijes/article/view/119251/112371

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39,* 83-101. https://doi.org/10.1017/S0261444806003399

Kepner, C. (1991). An experiment in the relationship of types of written feedback to the development of Second language writing skills. *Modern Language Journal, 75*, 305-313. doi:10.2307/328724

Klein, W., & Dittmar, N. (1979). *Developing grammars: The acquisition of German syntax by foreign workers.* Berlin, Germany: Springer.

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd edition). Los Angeles, CA: Sage.

Lee, G., & Schallert, D. (2008). Constructing trust between teacher and students through feedback and revision cycles in an EFL writing classroom. *Written Communication, 25*(4), 506-537. https://doi.org/10.1177/0741088308322301

Liu, J., & Hansen, J. (2002). *Peer response in second language writing classrooms.* Ann Arbor, MI: The University of Michigan Press.

Mackey, A., & Gass, S. (2012). *Research methods in second language acquisition: A practical guide.* Chichester, United Kingdom: Wiley-Blackwell.

Matsumura, L. C., Pathey-Chaves, G. G., Valdes, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality and third grade students' revision in lower- and higher-achieving urban schools. *The Elementary School Journal, 103*(1), 3-25.

Meisel, J. (1983). Strategies of second language acquisition: More than one kind of simplification. In R. Andersen (Ed.), *Pidginization and creolization as language acquisition* (pp. 120-157). Rowley, MA: Newbury House.

Menard, S. (2002). *Longitudinal research.* Thousand Oaks, CA: Sage.

Neuendorf, K. A. (2011). *The content analysis guidebook* (2nd edition). Thousand Oaks, CA: Sage.

Nordrum, L., Evans, K., & Gustafsson, M. (2013). Comparing student learning experiences of in-text commentary and rubric-articulated feedback: Strategies for formative assessment.*Assessment and Evaluation in Higher Education, 38*(8), 919-940.

Norris, J. M., Plonsky, L., Ross, S. J., & Schoonen, R. (2015). Guidelines for reporting quantitative methods and results in primary research. *Language Learning, 65*, 470-476. doi:10.1111/lang.12104

Phatiki, A. (2014). *Experimental research methods in language learning.* London, United Kingdom: Bloomsbury.

Polio, C., Fleck, C., & Leder, N. (1998). "If I only had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing, 7*, 43-68. http://dx.doi.org/10.1016/S1060-3743(98)90005-4

Poverjuc, O., Brooks, V., & Wray, D. (2012). Using peer feedback in a Master's programme: A multiple case study. *Teaching in Higher Education, 17*(4), 465-477. http://dx.doi.org/10.1080/13562517.2011.641008

Riazantseva, A. (2012). Outcome measure of L2 writing as a mediator of the effects of written corrective feedback on the ability of students to write accurately. *System, 40,* 421-430. http://dx.doi.org/10.1016/j.system.2012.07.005

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on

EFL writing quality. *TESOL Quarterly, 20*, 83-93. doi:10.2307/3586390

Rosansky, E. (1976). Methods and morphemes in second language acquisition. *Language Learning, 26*, 409-425. doi:10.1111/j.1467-1770.1976.tb00284.x

Schmidt, R. (1983). Interaction, acculturation and the acquisition of communication competence. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and second language acquisition* (pp. 137-174). Rowley, MA: Newbury House.

Schumann, J. (1978). *The pidgnization process: A model for second language acquisition*. Rowley, MA: Newbury House.

Semke, H. (1984). Effects of the red pen. *Foreign Language Annals, 17*(3), 195-202. doi:10.1111/j.1944-9720.1984.tb01727.x

Seror, J. (2011). Alternative sources of feedback and second language writing development in university content courses. *The Canadian Journal of Applied Linguistics, 14*(1), 118-143.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23,* 103-110. https://doi.org/10.1177/003368292023001O7

Simpson, J. M. (2006). Feedback on writing: Changing EFL students' attitudes. *TESL Canada Journal, 24*(1), 96-112. http://dx.doi.org/10.18806/tesl.v24i1.30

Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies, 10(2),* 29-46.

Van Beuningen, C. (2010). Corrective feedback in L2 writing: Theoretical perspectives, empirical insights and future directions. *International Journal of English Studies, 10(2),* 1-27.

Vyatkina, N. (2010). The effectiveness of written corrective feedback in teaching beginning German. *Foreign Language Annals, 43*(4), 671-689. doi:10.1111/j.1944-9720.2010.01108.x

Wildemuth, B. M. (2016). Longitudinal studies. In B. M. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (pp. 71-80). Santa Barbara, CA: ABC-CLIO.