Journal of Academic Language & Learning
Vol. 11, No. 1, 2017, A125-A144. ISSN 1835-5196

AALL
Association for Academic Language and Learning

The impact of explicit instruction and corrective feedback on ESL postgraduate students' grammar in academic writing

Amanda Müller and Carolyn Gregoric

College of Nursing & Health Sciences, Flinders University, Adelaide South Australia 5001, Australia

Email: amanda.muller@flinders.edu.au and carolyn.gregoric@flinders.edu.au

David R. Rowland

Student Services, The University of Queensland, Brisbane, QLD 4072, Australia.

Email: d.rowland@uq.edu.au

(Received 12 August, 2016. Published online 20 July, 2017.)

This study investigated the effects of a pilot program of written corrective feedback, supported by explicit grammar instruction, on a range of common errors made in ESL postgraduate writing. It sought to quantify improvement through the reduction of error rates and error categories. This small-scale quasi-experimental study involved 15 ESL doctoral students: 7 in an intervention group and 8 in a semi-equivalent control group. Participants wrote a short essay at the start of the study and a second essay ten weeks later at the conclusion of the study. These essays were analysed for errors, and it was found from linear regression analyses, assuming a constant percentage change in error rate across the groups, that the intervention group produced on average 43% fewer errors at post-test while the control group's error rate decreased by an average of 12% at post-test. In terms of error categories present in students' writing, the intervention group had on average 2.4 fewer categories at post-test, while the control group had on average 0.5 more error categories. These results suggest that for ESL doctoral students in the health sciences, substantial gains can be made within a maximum of 16 hours of written language support, while without such support, little if any improvements are likely to occur. Hence, despite the small sample size and the limitations of quasi-experimental designs, this study found evidence that such a program may produce improvement in the accuracy of ESL postgraduates' grammar, and this encourages further investigation of similar programs.

Key Words: ESL; postgraduate; corrective feedback; explicit grammar instruction.

1. Introduction

Students studying at a postgraduate level need to demonstrate advanced communication in ways appropriate to their discipline. Success, therefore, in postgraduate study is highly dependent upon students' writing and grammar skills (Gunawardena, 2014). Language proficiency can be an obstacle for international higher degree research students studying in Australia (Winchester-Seeto, Homewood, Thogersen, Jacenyik-Trawoger, Manathunge, Reid, & Holbrook, 2014). The English entry requirement for postgraduates can be as low as an IELTS score of 6.0 or equivalent (i.e. upper intermediate-advanced English). However, IELTS 6.0 is a level at which the IELTS test makers recommend further English study for students studying for university degrees where English is the medium of instruction (International English Language Testing Sys-

tem, 2013, p. 13). Thus, upon commencing their studies, it is common to find the writing skills of international postgraduate students are below a standard acceptable for a thesis. From previous research, it is known that English as a second language (ESL) nursing students' learning at the undergraduate level requires supplementary language support (e.g., Glew, 2013; Hillege, Catterall, Beale, & Stewart, 2014), such as effective written feedback (Giles, Gilbert, & McNeill, 2014). While ESL undergraduates may have problems with academic writing, this issue is most visible in the postgraduate arena because of its strong emphasis on writing. Poor writing ability has a negative impact on university studies, with postgraduate students having to pay editors or spend inordinate amounts of time ineffectively revising their written output. In addition, there is a corresponding increase in workload for academic staff (Gunawardena, 2014; Müller, Arbon, & Gregoric, 2015). The feedback that students receive from academic staff, at least at the undergraduate level, may be of variable quality and, therefore, of limited assistance in improving writing skills (Hyland, 2013). There is also pressure on postgraduate students to publish in academic journals, which requires significantly more effort for those not writing in their first language (Bomar, 2014). Consequently, additional assistance to improve academic writing ability is required by many ESL postgraduate students.

Notwithstanding the difficulties academic supervisors across disciplines can have with reading ESL postgraduate writers' work (Bitchener & Basturkmen, 2006), the program reported here was created specifically in response to a need identified within a university's nursing school. School staff reported concern about what they perceived to be a linguistic problem (for some further insight and verification on this, see Müller, 2011), rather than a subject-based problem (i.e. a lack of ability to produce comprehensible error-free language rather than a lack of knowledge about nursing). These postgraduate students, and their non-language specialist supervisors, needed a method for addressing the problem that would be attainable and sustainable. Thus, a program of language development that complemented currently available university support was designed, one which focused specifically on grammar-based proof-reading and error correction. The study reported here was designed to test the effectiveness of this program.

2. Research on explicit instruction and written correction feedback

According to Ferris (2011), 'errors are morphological, syntactic, and lexical forms that deviate from rules of the target language, violating the expectations of literate adult native speakers' (p. 3). Error correction assists students at all levels to recognise grammatical errors: error correction, and subsequent student modification of a text, is therefore a tool for learning and a means of preventing misunderstanding. Before proceeding, it should be noted that some errors in student writing are deemed 'untreatable' by this approach. Untreatable errors, such as inappropriate word choice, are those that cannot be merely resolved by applying rules (Ferris, 2011, p. 36).

As students advance in their English language learning, ongoing explicit grammar instruction may be of benefit for learning new structures and reinforcing known structures (Ellis, 2006, p. 102). Formal grammar instruction, particularly with attention to improving sentence composition, is appropriate to ESL thesis writers (Ferris, 2009; Gunawardena, 2014, p. A114). An important accompaniment to explicit grammar instruction is appropriate feedback. Written forms of corrective feedback (CF), also known as error/grammar correction (e.g., Truscott, 1999), can support ESL writers' self-mastery over linguistic errors by providing 'an indication to the learners that his or her use of the target language is incorrect' (Lightbown & Spada, 1999, p. 172). Attempts to increase the grammatical capacity of writers via CF are likely to be useful, providing that it is at the correct developmental level for the learner (Guénette, 2007, pp. 51-52) and that the learner pays attention to the feedback (Polio, 2012). While many studies have also found corrective feedback to be effective (e.g., Bitchener, 2008; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, Wright, & Moldawa, 2009), some studies are more critical, showing a negative effect on writing (e.g., Sheppard, 1992; Truscott, 2007). However, after reviewing the debate surrounding CF, Bitchener and Ferris (2012, p. 96) conclude that 'written CF clearly helps students to revise and edit their texts more successfully'.

CF can be provided through a variety of means. Feedback may not necessarily involve highlighting every error in a written text; instead, the feedback may draw attention to particular cat-

egories of errors only. Ellis (2009) identifies five strategies for addressing written CF: direct correction, indirect correction, metalinguistic correction, focused and unfocused correction, and electronic feedback. Of particular interest to this study is indirect metalinguistic-focused CF which indicates the location and type of error but the correct form is not automatically given to the writer. In the first instance, the program used in this study emphasised indirect error correction. However, if there were ongoing misunderstandings, the teacher then offered direct correction and further explanation.

Central to the strategy of providing indirect metalinguistic corrections in the program was the use of error codes (see Table 1), since these codes helped focus students' attention on the variety of errors possible and allowed the number of errors made in the students' writing samples to be quantified. The error categories given in Table 1 are frequently found in second language writing studies (a summary of the variations can be found in Bitchener & Ferris, 2012, p. 97) and are thus considered applicable to the participants. The use of coding, or cuing, is an indirect correction strategy because it indicates the error location without providing the corrected form. As a strategy for long-term improvement, indirect feedback strategies are preferred by students (Bitchener & Ferris, 2012, p. 94). Indeed, indirect teacher error correction – where the student is not told explicitly what the correct form is – may be more successful in the long term than direct error correction (Ferris, 2006, pp. 81-104); however, other research has also found that both the coded and uncoded forms of indirect CF are effective in the long-term improvement of composition writing (Bitchener & Ferris, 2012, p. 83). Providing advanced L2 writers with explicit written CF may reduce targeted linguistic errors by reminding them of grammatical rules (Bitchener & Knoch, 2010b). However, as Ferris, Liu, Sinha, and Senna (2013, p. 324) suggest, students' prior grammatical knowledge may be inadequate for their current level of study. This is a reason why some explicit grammar instruction was also incorporated into the teaching program.

Table 1. Error categories used for the program.

Code	Error Category	Examples
v	verbs (form/tense)	wrong verb or missing the verb, e.g. 'now it changed'
n	nouns	e.g. 'ability', 'car', 'advice', etc.
adj	adjectives	e.g. 'everyday', 'painful', 'expensive', etc.
adv	adverbs	e.g. 'quite', 'soon', 'however', etc.
prep	prepositions	e.g. 'in', 'at', 'on', 'for', 'with', 'by', 'to', etc.
art	articles	e.g. 'the', 'a', 'an', Ø
pl	plurals	e.g. car/cars, staff/staff, etc.
conj	conjunctions	e.g. 'and', 'but', 'nor', 'yet', etc.
punc	punctuation	e.g. , . : ; ' - etc.
s/v	subject-verb agreement	e.g. 'a person has', 'people have', etc.
sent	sentence fragment/run on	e.g. 'Having looked at the problem.'

Overall, the existing research indicates that the practice of written CF is effective in improving the linguistic accuracy of ESL writing. Much of this recent research relies on studies focused on one or several specific grammatical problems (e.g., Bitchener & Knoch, 2009b, 2010a, 2010b). The extent to which programs incorporating CF provided to postgraduate students are effective across multiple error categories has yet to be fully explored. Indeed:

> Future research on WCF needs to be conducted in authentic classrooms so that the feedback is given within the context of an instructional program, with ecologically valid writing tasks, and where revision is meaningful for the students … such studies need to be longitudinal …. [and] for feedback to

> be effective, it also needs to reflect and reinforce what is taught and emphasized in the class (Storch, 2010, p. 43).

Lee (2013) also calls for 'naturalistic studies in real classroom contexts' (p. 117) which further highlights the need for more classroom-based studies such as the current investigation.

Using evidence gathered from the literature review, a pilot program which combined teacher-led instruction and CF, mostly in the form of written correction was developed. The focus of the program would meet the needs of the school's staff and doctoral students and, at the same time, address a gap in the second language support offered to postgraduate university students generally. To explain, although a large number of pre-existing courses and resources at Flinders University deal with the style and structuring of academic text, they do not specifically focus on the range of grammatical issues found in postgraduate ESL writing. Thus, it was necessary to develop a new program of written language development for this group of ESL students. The idea behind the teaching program was to focus on error identification and corrective feedback, encourage students to help each other in this process (in lieu of their supervisor providing language help), and increase overall knowledge, awareness, and practice in proofreading generally. In designing the program, the researchers hypothesised that as a result of the program: (1) the overall combined error counts would be reduced; (2) the number of errors in each error category would be reduced; and (3) some error categories would be extinguished in the intervention group, but that little if any improvement would be seen in a control group. It was also expected that students would find that giving peer-based corrective feedback would improve their own awareness of common errors and increase their ability to produce error-free writing (although this specific issue is the subject of another paper).

The aim of this study was to investigate the effectiveness of the program on postgraduate writing. We sought to test how well written grammar could be developed by this program within a specific timeframe, and to quantify improvement as a function of error rate and error category. The research question was: to what extent can the program improve postgraduate students' grammatical accuracy, both in reducing the overall error rate and the number of error categories?

3. Material and methods

The study analysed pre- and post-test error rates and types found in the writing of 15 postgraduate students. Permission to conduct this study was gained from the Flinders University Social and Behavioural Research Ethics Committee (#6469).

3.1. Participants

All participants in the study were volunteer postgraduate students with ESL who originated from overseas (within the last ten years or less) and were undertaking a higher degree thesis in a health faculty at one Australian university. There was no coursework requirement for students. At the time of this study, all students were pursuing their own independent research under academic supervision. The students were upper-intermediate/advanced English users, since they had all passed the English entry requirement, which is an overall IELTS score of 6.0 with a minimum of 6.0 in each sub-test, or they came through an equivalent language pathway. Participants volunteering from one school in the faculty undertook the teaching program and participants from other schools in the faculty acted as the semi-equivalent control. The group acting as the control were offered the same program after the study had concluded, but received no feedback or instruction from the researchers prior to this. The intervention group consisted of seven students undertaking postgraduate nursing studies while the control group consisted of eight students undertaking non-nursing postgraduate studies. These were intact groups which the researchers could not randomly reassign. The researchers acknowledge that they were unable to control for the many factors outside of this program which may have resulted in the two groups being inequivalent on factors which might affect the results (e.g. the possibility that one school attracts more academically gifted students than the other).

3.2. Procedure

Prior to the commencement of the program, all participants handwrote and self-corrected a short essay in response to a prompt question. Given the diversity of students' research interests and to remove differences in content knowledge as a variable, a discipline-specific prompt was not used. Instead, the pre-test question was, 'University study is easy. Do you agree or disagree? Give your reasoning and examples.' Participants were given 40 minutes to complete the task, and at 30 minutes, an oral suggestion was given to spend the final 10 minutes correcting any errors. This procedure was repeated for the second essay at the end of the study, using a different general prompt question. The post-test question was, 'Living in another country is easy. Do you agree or disagree? Give your reasoning and examples.' The intervention group participants were given a copy of their corrected essays before the start of the program and after the post-test, while the control group received both corrected essays once the study had concluded.

3.3. Intervention Program

The program was conducted in two-hour blocks over eight weeks. It comprised focused grammar and error instruction and short practice sessions in the first hour, and focused indirect error feedback by peers and teachers on a thesis chapter draft in the second hour. Although we assumed that students had prior knowledge of grammatical forms, the focused grammar instruction brought any pre-existing discrepancies to light. The topics in the program were sequenced as shown in Table 2.

Table 2. Weekly program.

Week	Topics
Week 1	Introduction to underlining and coding for CF, the sentence and its variations, including clauses and phrases, how to identify the basic constituents of the sentence, and the relationship of punctuation to this.
Week 2	Verbs, including tenses, verb phrases, and their role in independent clauses.
Week 3	The use of articles and nouns, their function in noun phrases and role as the subject or object in sentences, plus plurals.
Week 4	Prepositions, phrasal verbs, and prepositional phrases.
Week 5	Phrases and clauses, including more work on verb, noun, and preposition phrases.
Week 6	Combining phrases and clauses to create different sentence types.
Week 7	Consolidation 1: supervised sessions correcting one error category at a time.
Week 8	Consolidation 2: supervised sessions correcting one error category at a time.

This overall approach of concentrating on one aspect of written language development in each instructional session is supported by prior research which indicates the effectiveness of focusing on one error category at a time, especially for discrete rule-based errors (Bitchener & Ferris, 2012, pp. 57-61). Furthermore, 'for advanced L2 writers, it is clear that one treatment on one error category can help them improve the accuracy of their writing' (Bitchener & Knoch, 2010b, p. 215).

In the first hour of each weekly lesson, the typical lesson format was:

1. Introduce the problem area, including where the related errors occur and providing contextual information about why this is an issue.
2. Provide explicit instruction about the error category, its related grammatical concepts, and show examples.
3. Identify resources that help students understand the error category, the grammatical and stylistic rules, and practical lists that teach strategic knowledge about how to gain help.

4. Provide explicit instruction on how the error category might be related to other error categories as part of the exploration of the issue.
5. Conduct a structured input activity asking students to identify and correct simplified forms of the error category in exercises provided by the teacher (some indirect correction may be needed at this point, with direct correction and restatement of the grammatical points given if there is still a problem).
6. Conduct a structured output activity asking students to produce their own sentences which do not make that particular error.
7. Ask students to swap their written work with other students and identify (but not fix) areas for correction (indirect correction) and the writer makes changes which the teacher later checks (direct correction).

Each focused instruction session was immediately followed by an hour-long hands-on workshop for focused indirect written corrective feedback which was conducted by peers in pairs, looking at each other's writing (often a few pages from a draft of their thesis).

The typical format of the second hour of the session was:

1. Students swap a piece of their writing (e.g. a page from a draft of their chapter) and correct the focal error category in their peer's work. Since one error category may overlap with another error category, some discussion of this relationship naturally emerged, but the focus needed to remain on the primary error category.
2. The teacher answers queries. This is where a discussion of the relationship between different error categories often arose, and was worthy of discussion without losing focus on the primary error category.
3. The teacher checks the corrections made by either the writer or the peer proofreader.

These workshops used focused corrective feedback because they emphasised that particular week's error category/focus ahead of other errors. Students were encouraged to provide a metalinguistic explanation of errors found to each other. Often, these peers attempted to provide some direct correction as well. By responding to input and processing the writing of others using given codes to help them focus, students were required to notice and understand the feedback. This was expected to support learning because the students had to internalise and consolidate new knowledge within their existing linguistic repertoire. The interests of this investigation are in the overall effects of the entire program (i.e. instruction and peer correction) rather than the effectiveness of the individual components.

3.4. Method of analysis

The essays were transcribed into electronic form, and independently coded for errors. This was achieved through blind coding of the essays, and on the few occasions where one coder missed an error, or where there was disagreement on the exact category of an error identified by both coders, a consensus was met before the final count profile was reached. To explain, the measurement method only counted one error for each point of difficulty, that is, something counted as a verb error may also have had a subject-verb and/or a sentence fragment error that was not coded. Joint decisions by the two authors about the error category to assign were based on the grammatical role the corrected version would play in the sentence. As outlined in Table 1, the error codes used were: verb form/tense, nouns, adjectives, adverbs, prepositions, articles, plurals, conjunctions, punctuation, subject-verb agreement and sentences (e.g. fragments and run-on sentences). These error categories were also the subject of instruction in the program. Student scores were entered into an Excel spreadsheet. The number of errors per 1000 words was calculated for each student for each essay, to avoid variations in essay length affecting scores. The base rate of 1000 words was selected because most academic written word counts are measured as multipliers of a thousand words, and this makes changes in error rates more tangible for the lay person. Analyses of the pre- and post-test essays were used to measure changes in students' writing ability, and a change statistic was calculated from errors per 1000 words in the pre-test and post-test (by subtracting the pre-test errors per 1000 words from the post-test errors per 1000 words such that a negative score indicated a reduction in errors, a score close to zero

indicated little change and a positive score would indicate an increased, rather than a decreased, number of errors at post-test). These change scores were used as the outcome variables for the analyses.

4. Results and discussion

4.1. Analysis of gross error rates

4.1.1. Preliminary analyses

Basic summary statistics for the intervention are shown in Table 3. Note first from this table that the error analysis was based on samples of around 300 words, which was considered adequate for an investigation of gross error rates because this is the minimum number of words required by the IELTS writing test for their essay writing task. However, samples of writing of this size may limit the analyses that can be done of error rates in the different categories given in Table 1 simply because so few errors were made in some categories (see Table 6 and related discussion).

Note also from Table 3 that the intervention group experienced a substantial drop in the average number of errors made while the control group did not, though because the average pre-test error rate for the intervention group was much higher than the average pre-test error rate for the control group, both groups coincidentally ended up with the same average post-test error rate. Following Lakens (2017), the statistical and practical significance of these changes was explored by plotting the mean difference score for each group, together with their associated 95% confidence intervals (CIs)[1] and estimates of educationally meaningful standardised gains (see Figure 1). Regarding what might constitute an educationally meaningful change, Wolf (1986) suggests that in an educational setting, a standardised gain[2] of $d = 0.25$ should be considered significant (something was learned) and $d = 0.50$ is practically/clinically significant (something really changed), and these gains (reductions of errors in this case) are indicated in Figure 1.

Table 3. Analysis of gross error rate (normalised to number of errors per 1000 words) by group. (M = mean, SD = standard deviation, M_d = mean of paired differences, CI = confidence interval.)

Group	Number of words in pre-test samples	Number of words in post-test samples	Pre-test error rate	Post-test error rate	Paired differences in error rates	
	M [Range]	*M* [Range]	*M* (*SD*) [Range]	*M* (*SD*) [Range]	$M_d(SD)$	95% CI
Intervention (n = 7)	316 [223, 402]	277 [240, 345]	114 (51) [58, 202]	65 (48) [22, 160]	-49.7 (42.0)	[-89, -11]
Control (n = 8)	307 [235, 407]	341 [228, 530]	74 (24) [33, 106]	65 (33) [19, 125]	-9.0 (23.6)	[-29, 11]

One key conclusion that can be drawn from Figure 1 is that the 95% CI for the change in error rate of the control group is consistent with both a conclusion that this group made no gains across the study period (zero is included in the interval and so the gain was not statistically significant[3]), and possibly that the control group students made educationally meaningful gains (d = –0.5 is included in the 95% CI). Such mutually inconsistent results mean that the sample size

[1] Confidence intervals provide a measure of the amount of statistical uncertainty in a result. See for example, Cumming and Fidler (2009) and Cumming and Finch (2005) for more details.

[2] That is, the difference in means divided by an appropriate standard deviation.

[3] A two-tailed dependent samples *t*-test yielded $p = 0.316$ (ns), but note that many authors have argued that the null hypothesis statistical test approach to analysing data is overly simplistic (see for example, Cumming, 2014), and this is why Figure 1 has been included.

had weak statistical power for detecting differences, and it should therefore be concluded that the data are inconclusive about what sort of gains the control group may have made (see for example, Lakens, 2017). In contrast, the 95% CI for the intervention group suggests that we can be much more confident that both a statistically significant[4,5] and educationally meaningful learning gain was made by the intervention group across the two months of the intervention, though again the small sample size has led to considerable uncertainty in the actual size of the gain.

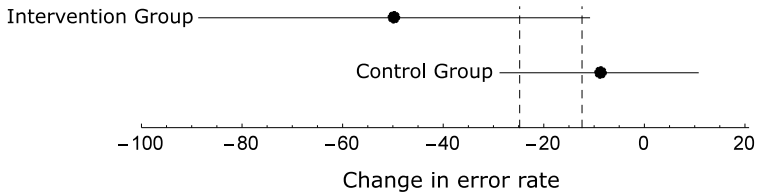

Figure 1. Estimated changes in gross error rates for the two groups (dots), together with the associated 95% CIs. The vertical dashed lines represent standardised learning gains of d = –0.5 and –0.25 respectively (reading from left to right), with the standard deviation of the error rate at the first measure of the intervention group being used as the standardiser.

The next question is how much of the gain achieved by the intervention group can be ascribed to the intervention, and how much might have occurred anyway due to the influence of other factors. It was for this purpose that a control group was included in the study. However, Figure 2 suggests that it would be a mistake to make a *direct* comparison in the mean changes in error rates for each group because the regression curves appear to diverge from the no change line and the two groups had considerably different initial mean error rates at pre-test.[6] That is, the size of the average learning gain for the intervention group would appear to not only depend on the intervention, but also on the initial average number of errors to start with: more errorful students have greater room to improve than less errorful students. To "control" for this initial difference, Figure 2 suggests that a regression analysis might be of value, and this is in fact commonly done (Scher, Kisker, & Dynarski, 2015). Consequently, the next sub-section discusses what can be learned from a regression analysis in this context.

[4] A one-tailed dependent samples *t*-test yielded $p = 0.010$.

[5] Note that a dependent sample *t*-tests assumes that the difference scores (i.e. the set of post-test minus pre-test error rates for each student) are approximately normally distributed. Tests of normality for the difference scores for the intervention and control groups yielded $p = 0.83$ and $p = 0.59$ respectively, so the use of *t*-tests is justified. However, it cannot be assumed that this result would necessarily hold for larger samples as it has to be acknowledged that the results from small sample sizes are subject to more potential random variation than are large samples.

[6] Although a two-tailed independent samples *t*-test revealed that the pre-test means were not statistically significantly different ($p = 0.088$), the standardised difference in the initial mean error rates is quite high, being 1.67 or 0.799, depending on whether one uses the standard deviation of the initial error rates of the control or intervention groups respectively as the standardiser. That these large standardised differences were not found to be statistically significantly different from zero could mean one of two things. First, it could be that the two groups are in fact samples from the same distribution of abilities, but because the variation in abilities seen in the general population is quite large, these small samples by chance came from different ends of the total distribution. Another possibility suggested by a consideration of the relative distributions of initial error rates shown in Figure 1 is that three members of the intervention group had unusually high error rates to start with for some reason, and this is what led the two groups to be substantially different on average at baseline. More data will be needed to resolve this issue.

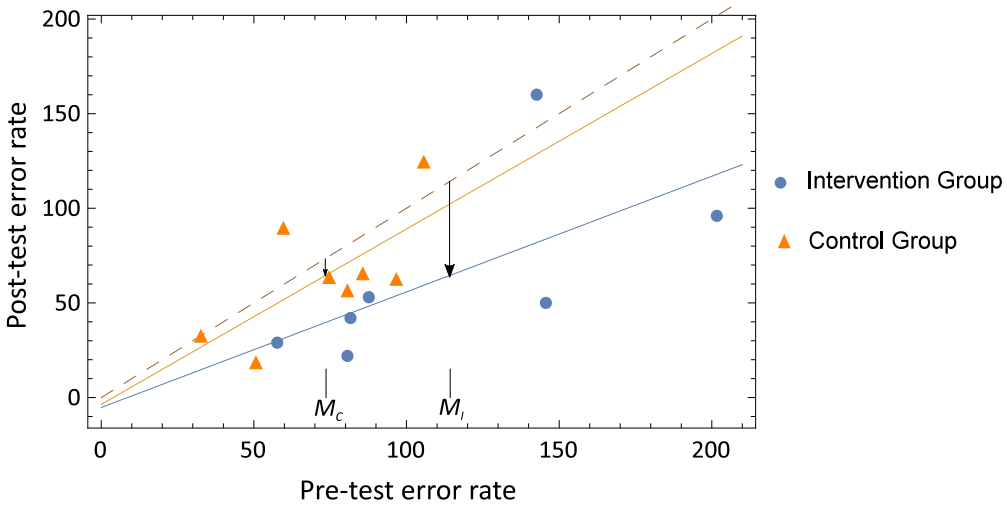

Figure 2. Comparison of the gross error rates (normalised to number of errors per 1000 words) at the pre- and post-tests for the control group (triangles) and intervention group (circles). The two solid lines represent the unconstrained[7] ordinary least squares regression lines through each scatter plot, while for comparison, the dashed line indicates points for which there has been no change in error rate between the two tests. The two small vertical lines labelled M_C and M_I indicate the mean error rates at the pre-test for the control and intervention groups respectively. The lengths of the associated downwards pointing arrows thus indicate the mean reduction in gross error rate for each group over the two months of the study.

4.1.2. Linear regression analyses

As this study used a quasi-experimental design, before proceeding with a regression analysis the question of the comparability of the intervention and control groups needs to be revisited as the more covariates which might influence the results on which the two groups are equivalent, the more confident one can be in the findings (Scher et al., 2015; Shadish & Cook, 2009). In relation to this issue, recall from Sub-Section 3.1 that the two groups were comprised of students who could be expected to have similar levels of ELP, were at the same level of study (PhD), and in related fields of study (health sciences). They were also at similar stages in their studies (towards the end of first year or in their second year) and the slight differences in their stage of study were not expected to be an important factor influencing ELP at the start of the study. Support for this expectation comes Craven (2012, p. 11), who found that that over an academic degree (unlike the 8 weeks of this study), no student in the study group was able to improve more than one IELTS band and 35% did not improve or even worsened by half an IELTS band. In addition, written grammatical skill has also been shown not to improve significantly over a degree either (Knoch, Rouhshad, Oon, & Storch, 2015). Figure 2 also reveals a considerable overlap in initial error rates between the two groups, even though some more extreme results in the intervention group caused the starting mean error rates to be considerably different. Consequently, apart from some differences in initial error rates, it seems reasonable to consider the two groups sufficiently comparable for comparisons to be reasonably valid.

Another issue to consider, however, before attempting to control for initial group differences with a regression analysis is how much "extrapolation" will be needed, as with regressions, the greater the extrapolation the greater the potential error. Because of this issue, Scher et al. (2015, p. 6) recommend that regression techniques only be used to control for initial quasi-experimental standardised differences of less than 0.25 standard deviations. As the two groups in this study vary initially by considerably more than this, a direct statistical comparison of the

[7] That is, the regression lines were allowed to have non-zero intercepts, in contrast to the approach taken in constructing Figure 3.

two groups using regression techniques was deemed to be infeasible.[8] However, in order to lay a foundation for subsequent studies, and because of added insights one can obtain into each group alone, a partial analysis using regression techniques was still conducted, and the results will be presented next.

If a simple linear regression model is applicable in this case, then there are three possibilities: (a) an on average uniform learning gain (i.e. reduction in error rate) across the intervention group; (b) a uniform *percentage* decrease in error rate; and (c) a model intermediate between the first two: that is, that there is a linear reduction in error rate across the group, but it varies with initial error rate (i.e. there is an interaction term in the linear regression model). Although there is, in fact, insufficient data on which to determine which of these models is best, for reasons given in Appendix B, model (b) was chosen for the analysis. That is, it was assumed that for each group,

$$(\text{Change in Gross Error Rate})_i = b_1 \times (\text{Initial Gross Error Rate})_i + e_i,$$

where the subscript *i* refers to the *i*th participant in the group, $b_1 \times 100\%$ gives the average percentage reduction in gross error rate for the group in question, and e_i is a random error term (i.e. it accounts for other random factors which may affect students' changes in error rate, such as seeking additional help or just having an 'off day' when they re-sit the test).

Generally, when one wants to use regression techniques to control for initial comparison group differences, one uses a multiple regression approach, and specifically an ANCOVA (Weber, 2009), and this would be the approach that would be taken here if the first model mentioned above was used. However, with forced regression through the origin, the best fit lines through each group are determined independently of each other, and with an assumed constant percentage change across each group, an offset in initial average error rates does not influence the conclusions. Consequently, regression analyses for the intervention and control groups were done independently, with the results being shown in Table 4 and Figure 3. The results are thus consistent with a conclusion that the control group likely did not improve to any great degree over the study period, but that the intervention group did. Hence, it would seem that much, if not all, of the improvement in the intervention group can be ascribed to the impacts of the intervention, though with such a small amount of data there is a considerable amount of uncertainty in the exact figures.

Table 4. Predicted percentage changes in error rates of the regression model.[9]

Group	% change in gross error rate per 1000 words	95% Confidence Interval	p[a]
Intervention	–43	[–71, –15]	0.009
Control	–12	[–37, 14]	0.311

[a] Two-tailed *t*-test of the hypothesis that $b_1 = 0$.

There is one other issue to discuss before leaving this section, and that is the possibility that there is an outlier in the intervention group's data, namely the one student who did not show a substantial reduction in gross error rate at post-test (see the one blue dot in Figure 2 which lies above the dashed line). This student was identified early in the program as not having the commencing English skill level expected of a postgraduate student, and it was expected that the activities would be too advanced to be of benefit for this student. Consequently, it is possible that this student's results are an outlier, inconsistent with the general trend. However, as the sample

[8] The preceding comments apply to the usual case of regression curves which are not constrained to go through the origin. We will, however, later conclude that a regression through the origin approach is the best approach for this study. Nevertheless, similar conclusions are expected to hold in that case as well.

[9] Calculated using *Mathematica 10.3*'s LinearModelFit algorithms.

size is too small to draw this conclusion with any certainty, this student's results have not been removed from the analysis. To provide a foundation for future studies though, the Appendix provides a discussion of statistical methods which could be used to establish whether or not the noted student might be an outlier.

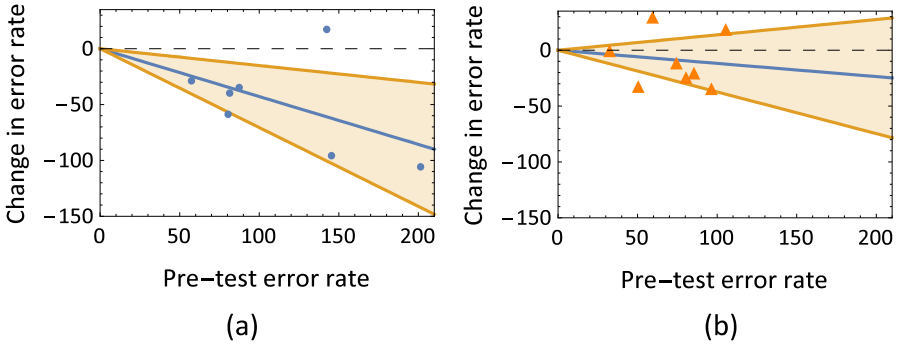

Figure 3. Change in error rate versus pre-test error rate for (a) the intervention group and (b) the control group. The dots/triangles are the data points, the blue line is the best fit linear regression line constrained to go through the origin, the orange lines give the corresponding 95% prediction limits on the best fit lines, and the horizontal dashed line is the 'no change in error rate' line.

4.2. Error category

Apart from gross error rate, it is also of interest to consider the impact of the intervention on error rates in the particular grammatical categories given in Table 1. In particular, it was of interest to see if any categories became extinct as greater mastery was achieved. The results are shown in Figure 4 and Table 5, but it should be noted that the number of errors in some categories (i.e. the ones below the break in the middle of Table 5) were too small for conclusions to be drawn with any confidence. However, consistent with the conclusions made above that the intervention group appears to have made considerable reductions in error rates while the control group are unlikely to have done so, Figure 4 shows that all but one member of the intervention group decreased the number of error categories occurring in their writing (mean difference in error rate categories M_d = –2.4, SD = 1.6), while for the control group, about half the group saw an increase while the other half saw a decrease or no change (mean difference in error rate categories M_d = +0.5, SD = 2.0). This result for the control group suggests that these changes are just random fluctuations indicative of either uneven control of written English grammar, or indicative of the sort of natural variation in the number of errors one is likely to see in different samples of writing.

The results in Table 5 also show why it is problematic to compare the changes in error rates between the intervention and control groups by error category. With zero or close to zero errors initially in categories such as subject-verb agreement, adverbs, conjunctions, and sentence fragments/run-ons, the control group had little if any room to improve in these categories. Consequently, comparisons of gains made in a group where gains can be made with a group where next to no gains are possible, are not valid. In addition, in categories such as articles, verbs and prepositions, the average initial error rate for the intervention group was around double the average initial error rate for the control group. Thus, if it is true that some errors in some categories are easier to address than others (recall from Figure 1 that the gross error rate reduction has been assumed to grow linearly with the size of the initial gross error rate), then such errors are more likely to be found in the intervention group than the control group, giving them an "unfair advantage" in an error reduction comparison.

Despite these limitations, Table 5 is of interest in identifying the most common error categories (articles, verb form/tense, prepositions, punctuation and plurals), and that with the intervention group, substantial reductions in error rates were achieved in the categories of articles, verb

form/tense, prepositions and punctuation. In addition, although the initial error rates were low, it appears that issues with subject-verb agreement and adverbs may have come close to being nullified with the intervention group.

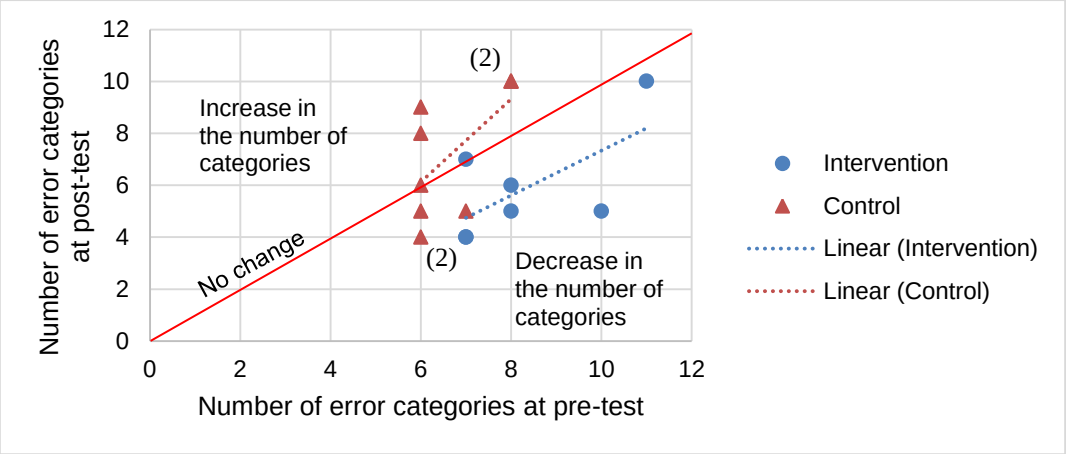

Figure 4. Change in the number of error categories from pre-test to post-test. The dotted lines are the linear regressions through the scatterplots for each group, while the solid diagonal line indicates no change from pre-test to post-test. The vertical distance from the 'no change' line to a data point gives the change in the number of error categories seen in that student's writing. The numbers in brackets indicate the number of students having the same set of results.

Table 5. Error rate improvement per 1000 words, by error category. Note that below the midline, there was an insufficient number of errors made for conclusions to be drawn with any confidence.

	Errors per 1000 words of Intervention Group			Errors per 1000 words of Control Group		
Error Type	Pre-test *M* (*SD*) [Range]	Post-test *M* (*SD*) [Range]	Change M_d (*SD*) [Range]	Pre-test *M* (*SD*) [Range]	Post-test *M* (*SD*) [Range]	Change M_d (*SD*) [Range]
Articles	29.6 (14.8) [10.6, 49.8]	11.1 (8.6) [0, 20.8]	-18.6 (8.6) [-29.0, -9.2]	14.1 (6.2) [8.2, 22.4]	14.0 (10.0) [0, 24.8]	-0.1 (11.9) [-18.7, 16.6]
Verbs (form/tense)	22.1 (13.5) [10.6, 49.3]	13.4 (9.5) [3.7, 26.0]	-8.7 (10.4) [-24.3, 7.0]	13.1 (8.0) [0, 25]	9.8 (11.0) [0, 27.2]	-3.4 (7.6) [-11.2, 7.1]
Prepositions	19.1 (20.0) [2.5, 58.3]	11.7 (13.0) [0, 33.5]	-7.4 (12.3) [-33.3, 1.1]	11.9 [6.7) [3.6, 22.1]	8.6 (10.7) [0, 32.6]	-3.2 (12.3) [-17.9, 21]
Punctuation	14.3 (6.0) [7.6, 24.9]	5.3 (4.3) [0, 11.2]	-9.1 (5.7) [-16.6, -0.3]	15.2 (10.9) [2.7, 27.0]	9.5 (4.7) [0, 14.2]	-5.7 (12.4) [-22.4, 10.7]
Plurals	11.1 (2.9) [7.1, 15.2]	12.1 (13.1) [0, 33.5]	1.0 (12.9) [-15.2, 22.5]	7.5 (4.5) [0, 12.8]	6.1 (5.2) [0, 13.6]	-1.4 (4.4) [-8.4, 5.0]

Table 5 continued

Error Type	Errors per 1000 words of Intervention Group			Errors per 1000 words of Control Group		
	Pre-test *M* (*SD*) [Range]	Post-test *M* (*SD*) [Range]	Change M_d (*SD*) [Range]	Pre-test *M* (*SD*) [Range]	Post-test *M* (*SD*) [Range]	Change M_d (*SD*) [Range]
Adjectives	4.6 (5.4) [0, 14.2]	2.6 (4.1) [0, 11.2]	-2.0 (6.6) [-14.2, 7.5]	4.1 (7.0) [0, 18.7]	3.0 (2.8) [0, 7.1]	-1.1 (7.2) [-7.1, 15.2]
Subject-verb agreement	3.7 (3.1) [0, 7.7]	0 (0) [0, 0]	-3.7 (3.1) [-7.7, 0]	0 (0) [0, 0]	1.98 (4.61) [0, 13.2]	1.98 (4.61) [0, 13.2]
Adverbs	3.5 (2.2) [0, 7.3]	1.1 (1.9) [0, 4.2]	-2.3 (1.9) [-4.5, 0.6]	0.5 (1.2) [0, 3.6]	0.9 (1.8) [0, 17.5]	2.6 (6.3) [-3.6, 17.5]
Nouns	3.0 (5.4) [0, 14.7]	3.6 (5.3) [0, 14.9]	0.7 (3.0) -3.6, 4.2]	3.6 (3.6) [0, 8.6]	1.9 (2.1) [0, 4.5]	-1.7 (4.9) [-7.3, 4.5]
sentence fragment/run on	2.0 (2.0) [0, 4.5]	2.1 (4.2) [0, 11.2]	0.05 (4.2) [-4.5, 7.5]	1.7 (2.7) [0, 7.5]	3.8 (2.9) [0, 8.1]	2.1 (3.6) [-4.0, 8.2]
Conjunctions	1.2 (2.2) [0, 5.1]	1.6 (2.9) [0, 7.4]	0.3 (2.9) [0, 7.4]	1.9 (3.1) [0, 8.6]	2.9 (3.4) [0, 8.1]	1.0 (2.4) [-0.4, 6.8]

4.3. Discussion

Our investigation sought to explore the extent to which a program of focused grammar and error instruction and practice, together with focused indirect error feedback by peers and teachers on a thesis chapter draft could improve postgraduate ESL students' overall error rate. The literature was unclear about the likely extent of improvement across a program that targeted a broad range of error types. The present study found that the writing skills of the students in the intervention group demonstrated a statistically significant improvement (43% fewer errors), whereas the changes in the control group (12% fewer errors) did not reach statistical significance. This finding for the intervention group is in contrast to a study of EFL graduate student writing by Truscott and Hsu (2008) which found that error correction (via underlining the location of errors) improved only the marked text, but not subsequent writing. Perhaps the difference in results between our study and Truscott and Hsu (2008) is attributable to differences in how errors were marked and the explanations given in our program (where errors were examined as a part of identifiable phrases or clauses). It might also have something to do with how the students were performing before the intervention – it is possible that a program of the type trialled does not work for those who have a level of English which is too low to access the program. Students cannot correct errors without the grammatical foundation to understand those errors. This conjecture also brings to mind the research which indicates that the type of instruction must be at the level of the learner (Ferris, 2009, pp. 101-124), and needs to avoid a mismatch between a student's developmental readiness and the feedback provided (Polio, 2012, p. 384).

Our investigation also sought to investigate to what extent the program could reduce the number of error categories appearing in postgraduate ESL students' writing. In relation to this question, overall, all but one of the participants in the intervention program had fewer error categories at post-test, whereas this was true for only three of the eight students in the control group (see Figure 4). These results suggest that teacher instruction, followed by peer and teacher CF, can effectively target a range of error categories over time, and this is especially true for rule-based

errors (verbs, articles) which seem more amenable to written CF (Bitchener & Ferris, 2012, p. 64; Ferris, 2011, p. 36).

More specifically, in regards to article use, the average 43% reduction in error rates for the intervention group (Table 4) corroborate a study by Bitchener and Knoch (2010b) on advanced L2 learners' articles. That study demonstrated the effectiveness of once-off provision of written CF, particularly direct CF with a written meta-linguistic explanation and example, in the treatment of article errors. Numerous other studies (e.g., Bitchener, 2008; Bitchener & Knoch, 2009a, 2009b; Bitchener & Knoch, 2010a; Bitchener, Young, & Cameron, 2005; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, Wright, & Moldawa, 2009) also report written CF to be effective in correcting article errors. However, there are other research papers which show poor responses to written CF. Ferris (2006), for example, reports an increase in the number of article errors made by undergraduate students, while a study of university students in Japan reports no improvement for indefinite articles (Shintani, Ellis, & Suzuki, 2014). Perhaps the program was effective in the current study because article use was addressed within the context of noun phrases and the management of plurals, and there was an emphasis on the use of a decision tree to decide on whether or not an article was needed (thus enforcing the rule-based nature of this aspect of grammar).

A similar point about giving CF and teaching the identification of errors in context can be made for verbs, which the program taught as positioned in verb phrases and indexed against a subject in a larger independent clause. From a relatively large initial error rate with verbs, the intervention group experienced a substantial average reduction of 39% (Table 5). Our findings are consistent with other research studies, such as Bitchener, Young and Cameron (2005), who report significantly increased accuracy in the use of the past simple tense by post-intermediate adult English learners when written CF is accompanied by conference feedback. Other CF studies report a reduction in verb errors made by undergraduate ESL students (Cowan, Choo, & Lee, 2014) and EFL university students (Meihami & Meihami, 2013).

What is perhaps more surprising in this study is that prepositions, a non-rule-based error, was improved by 7.4 errors per 1000 words in the intervention group. Bitchener, Young and Cameron (2005) also found that longer term improvement in the use of prepositions is more likely to be achieved when written CF is combined with conference feedback, even if it causes less accuracy in the short term. The program situated the preposition as either belonging to a verb, phrasal verb, preposition phrase, or idiom, and it is possible that this contextualisation contributed to this positive result.

5. Limitations

The current study has a number of limitations. First, even though the study sampled an intervention group of around half of the possible pool of ESL research students in the nursing school at the university, the sample size itself was small. This means that while the participating students may have been representative of the population from which they were sampled, there may not have been enough of them to fully capture the amount of variation seen in such students elsewhere. Consequently, while in the design of the study we assumed that parameters such as language ability (i.e. a minimum standard of 6.0 in all IELTS subtests) would be roughly equivalent for the two groups, in reality this appeared to be untrue. Since the groups were pre-existing, rather than randomly allocated, it appears that the control group started as a more competent writing group and had less room to improve than the intervention group. Furthermore, between-group differences and within-group variations can have an impact on the extent to which one intact group can be a control for the other. In short, it is unclear how important this group difference is and what impact it has on the findings. It would be interesting to establish whether there is an optimum skill level for benefiting from this particular program and how outcomes from the ideal program differ depending on students' starting skill level.

Second, and related to the previous point, it became apparent that the English skills of one intervention group participant were well below the standard usually required for entry into postgraduate study. The results of this participant bear out the notion that the program is perhaps not

appropriate for students with much lower language skills than intended. As a result, the participant became an outlier that reduced the estimated impact of the intervention. Nonetheless, as there was not enough data to definitively identify this student as an outlier, for reasons of accuracy and integrity in maintaining the variation in population parameters, this student's data were not removed.

6. Conclusion

International postgraduate students require support for their writing, and one way of enhancing writing ability is to provide explicit instruction in grammar accompanied by written CF, which was the focus of this investigation. We sought to test how well writing skills could be developed by this program within a specific timeframe, and to quantify improvement as a function of error rate and category. The research question asked about the extent to which the program could improve postgraduate students' grammatical accuracy. The short answer is 43% fewer errors and nearly 3 fewer error categories on average, gained within a maximum of 16 hours written language support. The results of this study indicate that the program had an effect on reducing both the number of overall errors and the number of error types found in students' work, while a semi-equivalent control group did not show a statistically significant improvement over the same time period in the absence of an intervention.

Overall, it appears that students participating in the program were by the end of the program, better able to identify and correct target errors than they were at the beginning. The program increased students' ability to notice and respond to errors. Moreover, feedback received, whether from the teacher or peers, was personalised to the needs of each participant. Throughout the program, students in the intervention group had multiple opportunities to develop their understanding of grammar rules including through meta-linguistic information provided by the teacher and student explanations to peers. Meaningful exposure and multiple practice opportunities avoided overwhelming students by not requiring them to simultaneously notice too many error categories. Thus, from this investigation it appears that the combination of instruction and feedback, both indirect and then direct, improved students' written accuracy more than they could be reasonably expected to achieve on their own. Although the design of this investigation did not include a delayed post-test, measurable performance improvements generally require longer time spans. A review of written CF studies by Bitchener (2012) reports that written CF promotes long-term retention, so it could be speculated that this would be also the case for our participants too.

To conclude, this investigation provides additional glimpses into how focused teacher instruction and CF, in combination, may improve the self-mastery of linguistic errors by ESL postgraduate students. From this investigation, it appears that the overall written accuracy and self-editing skills of postgraduate ESL students across multiple error categories may be improved through a program providing metalinguistic explanation via teacher instruction followed by an opportunity to apply this knowledge through the provision of CF to peers. The promising results achieved from this small pilot program suggest that further larger scale studies with equivalent control groups are warranted.

Acknowledgements

This study was funded by a start-up grant from the Flinders University School of Nursing & Midwifery. The authors would like to acknowledge the contribution of the anonymous reviewers' feedback on this paper. For details of the writing program itself or possible collaborations, please contact Dr Amanda Müller, who welcomes your involvement.

Appendix A. Testing for outliers in a regression analysis

The first assessment approach to assess for outliers is to determine whether the residuals are plausibly normally distributed about the regression lines using both Q-Q plots (e.g. Howell (2010, pp. 77-79)) and a normality test, as this is an assumption about the data made by regres-

sion formulas. With the potential outlier included in the analysis, the residuals failed a normality test[10] ($p = 0.027$[11]), while they did not when the potential outlier was excluded from the analysis ($p = 0.66$), with the Q-Q plots supporting this conclusion. Thus, in comparison to the rest of the data, the potential outlier student's data point is much further from the regression line than one would expect of a set of normally distributed residuals. However, while this result *may* mean this student is an outlier, it may also mean that the residuals are not normally distributed in this case. Substantially more data would be needed to determine which conclusion is correct.

Whether the "unusual" student's result should be of any concern, however, depends also on how much "influence" it has on the regression analysis as whenever one does a regression analysis, there is always the possibility that a point with high "leverage" (i.e. lying near either edge of the scatter plot) and high "distance" (i.e. the point lies a long way off the regression line) has an unduly high "influence" on the location of the regression line (see for example, Howell, 2010, pp. 541-542). From Figure 2, we can see that the potential outlier might fit these criteria. One measure of influence is Cook's *D*, and while there are varying guidelines as to what values of Cook's *D* should be of concern, Howell (2010, pp. 541,2) states that values over 1 are unusual, and the potential outlier has a Cook's $D = 0.96$ while all other points have Cook's *D*s under 0.25. Thus, there is reason to be concerned about the influence of this student's results on the analysis.

The above considerations suggest that the potential outlier possibly is an outlier that can be justifiably removed from the analysis. However, given that the above analyses are based on just seven data points, and the estimated percentage drop in gross error rate for the intervention group when the "outlier" is removed still lies well within the 95% confidence bands shown in Table 4 and Figure 3, no further consideration was given to this issue until more data can be obtained to reduce uncertainties.

Appendix B. Discussion of alternative linear regression models

As mentioned in Sub-Section 4.1.2, one possible linear regression model is that there was, on average, a uniform learning gain (i.e. reduction in error rate) across the intervention group. Such a model cannot be indefinitely valid though, as to take an extreme example to illustrate the point, a student with a zero error rate to start with cannot experience a drop in error rate as this model requires. However, it is possible that if the initial set of error rates are not too close to zero, that across that set there can be an approximately uniform drop in error rate but that at some point when the trend line gets closer to a zero rate, it makes a turn so that it ends up passing through the point (0, 0).

A second simple linear possibility is that on average, there were uniform *percentage* decreases in error rates across the groups. One reason to think that this is a possibility is that a uniform percentage decrease in error rate means that those students with the highest error rates to start with on average make the largest improvements, with those with lower error rates to begin with making proportionally lower gains on average. This is a plausible model as some errors are probably easier to correct than others, with students with a high error rate to start with more likely to have sets of such errors than those students who have low error rates to start with, and who have maybe started to plateau in their rates of improvement because their remaining types of error are more resistant to change. A second reason to prefer this model over the first one is that it automatically satisfies the theoretical requirement that the model should predict that an initially error-free student should also be error-free at post-test.

The third possibility is intermediate between the first two: there is a linear reduction in error rate across the group, but it varies with initial error rate (i.e. there is an interaction term in the linear regression model). As with the first model, this model will have a non-zero intercept and so at

[10] Using *Mathematica 10.3*'s standard DistributionFitTest function.

[11] $P < 0.05$ means the data are unlikely to be a random sample from a normal distribution; $p > 0.05$ means the data are not inconsistent with a random sample from a normal distribution.

some point when the initial error rate gets sufficiently close to zero, there must be a change in direction to the regression line so that it passes through the origin.

There is, in fact, insufficient data on which to determine which of the above models is best. However, there is a technical issue with the naïve application of the first (and third) model that researchers need to be aware of. This issue arises because a sample of only about 300 words provides only an *estimate* of a student's initial average error rate: different samples, even taken at the same time, will have different error rates, and this can be seen in the scatter about the regression line of the control group's data (see Figure 2), assuming that they did not in fact improve across the course of the study. As a result, the covariate on which the regression is made, namely the pre-test error rate, is "measured with error".[12] The consequence of this is a phenomenon known as "regression dilution" or "regression attenuation" which can result in erroneous conclusions being made about group differences if those groups have different initial mean error rates. See Trochim (2006, http://www.socialresearchmethods.net/kb/statnegd.php) for a clear diagrammatic explanation of this phenomenon.[13]

In the absence of any contrary evidence, of the above models the second model seems the most attractive (and it can be shown that regression dilution affects it to a far lesser degree than is the case for the first model), but as it forces the regression curve to go through the origin, it is recommended for statistical analysis reasons that such a model should only be used if one is quite confident that it applies (Eisenhauer, 2003). Eisenhauer (2003) suggests three criteria for deciding whether or not to use a regression through the origin model. First, the model should be a theoretical requirement, which is plausibly the case for this problem: a perfect, error-free student to start with should be error free also at post-test. Second, the unconstrained linear regressions through the data should come reasonably close to passing through the origin. That this is the case can be seen from Figure 1, which shows that the intercepts for the two unconstrained linear regressions through the pre-post error rates are both relatively small, and not statistically significantly different from zero, being –5.23 (p = 0.902) and −3.61 (p = 0.909) for the intervention and control groups respectively. The final criterion is that the resulting statistical model should provide a superior fit to the data and this is also the case. Evidence for this claim comes from the fact that models which were not constrained to go through (0,0) had confidence bands for the regression lines which predict that an initially error free student could have a large error rate at post-test, and negative error rates post-test were a predicted possibility even for moderately high initial error rates.

References

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102-118. doi: http://dx.doi.org/10.1016/j.jslw.2007.11.004

Bitchener, J. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing, 21*(4), 348-363. doi: http://dx.doi.org/10.1016/j.jslw.2012.09.006

Bitchener, J., & Basturkmen, H. (2006). Perceptions of the difficulties of postgraduate L2 thesis students writing the discussion section. *Journal of English for Academic Purposes, 5*(1), 4-18. doi: http://dx.doi.org/10.1016/j.jeap.2005.10.002

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

Bitchener, J., & Knoch, U. (2009a). The relative effectiveness of different types of direct written corrective feedback. *System, 37*(2), 322-329. doi: http://dx.doi.org/10.1016/j.system.2008.12.006

[12] This does not mean that the number of errors in a sample of writing cannot be measured exactly, only that the resulting number cannot be considered to be an exact measure of a student's average error rate at that point in time.

[13] One approach to correcting for this problem can be found on the cited webpage, but note that it requires a measure of the test-retest reliability of the writing test used to determine students' error rates.

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

Bitchener, J., & Knoch, U. (2010a). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31*(2), 193-214. doi: 10.1093/applin/amp016

Bitchener, J., & Knoch, U. (2010b). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*(4), 207-217. doi: http://dx.doi.org/10.1016/j.jslw.2010.10.002

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*(3), 191-205. doi: http://dx.doi.org/10.1016/j.jslw.2005.08.001

Bomar, P. J. (2014). How to create a 'writing tool box'for scholars writing in English as a second language. *International Journal of Nursing Practice*, *20*(S1), 1-8. doi: 10.1111/ijn.12270

Cohen, J. (1988). *Statistical power analysis for behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cowan, R., Choo, J., & Lee, G. S. (2014). ICALL for improving Korean L2 writers'ability to edit grammatical errors. *Language Learning & Technology, 18*(3), 193-207. Retrieved from http://llt.msu.edu/issues/october2014/cowanchoolee.pdf

Craven, E, 2012, 'The quest for IELTS 7.0: Investigating English language proficiency of international students in Australian universities', *IELTS Research Reports,*Volume 13 (pp. 1-61), Canberra: IELTS Australia/London: British Council.

Cumming, G. (2014). The New Statistics: Why and how. *Psychological Science, 25*(1), 7-29. doi:10.1177/0956797613504966

Cumming, G., & Fidler, F. (2009). Confidence intervals: Better answers to better questions. *Zeitschrift für Psychologie/Journal of Psychology, 217*(1), 15-26. doi:10.1027/0044-3409.217.1.15

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*(2), 170-180. doi:10.1037/0003-066X.60.2.170

Eisenhauer, J. G. (2003). Regression through the Origin. *Teaching Statistics, 25*(3), 76-80. doi:10.1111/1467-9639.00136

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 83-107. doi: 10.2307/40264512

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, *63*(2), 97-107.

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System, 36*(3), 353-371. doi: http://dx.doi.org/10.1016/j.system.2008.02.001

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). New York: Cambridge University Press. doi: https://doi.org/10.1017/CBO9781139524742

Ferris, D. (2009). *Teaching college writing to diverse student populations*. Michigan, USA: The University of Michigan Press. doi: 10.3998/mpub.263445

Ferris, D. (2011). *Treatment of error in second language student writing* (2nd ed.). Michigan, USA: The University of Michigan Press. doi: 10.3998/mpub.2173290

Ferris, D. R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing, 22*(3), 307-329. doi: http://dx.doi.org/10.1016/j.jslw.2012.09.009

Giles, T. M., Gilbert, S., & McNeill, L. (2014). Do students want and need written feedback on summative assignments? Engaging students with the feedback process-a topic review activity. *ergo*, *3*(1), 15-22. Retrieved from http://www.ojs.unisa.edu.au/index.php/ergo/article/view/938

Glew, P. J. (2013). Embedding international benchmarks of proficiency in English in undergraduate nursing programmes: Challenges and strategies in equipping culturally and linguistically diverse students with English as an additional language for nursing in Australia. *Collegian*, *20*(2), 101-108. doi: http://dx.doi.org/10.1016/j.colegn.2012.04.002

Gunawardena, M. (2014). Developing effective pedagogies of grammar: The two-in-one approach. Journal of Academic Language and Learning, 8(3), A113-A123.

Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, *16*(1), 40-53. doi: http://dx.doi.org/10.1016/j.jslw.2007.01.001

Hillege, S. P., Catterall, J., Beale, B. L., & Stewart, L. (2014). Discipline matters: Embedding academic literacies into an undergraduate nursing program. *Nurse Education in Practice*, *14*(6), 686-691. doi: 10.1016/j.nepr.2014.09.005

Howell, D. C. (2010). *Statistical methods for psychology* (7th ed. ed.). Belmont, CA: Belmont, CA : Thomson Wadsworth.

Hyland, K. (2013). Student perceptions of hidden messages in teacher written feedback. *Studies in Educational Evaluation*, *39*(3), 180-187. doi: http://dx.doi.org/10.1016/j.stueduc.2013.06.003

International English Language Testing System. (2013). *IELTS guide for educational institutions, governments, professional bodies and commercial organisations.* British Council/ IDP: IELTS Australia/Cambridge English Language Assessment. Retrieved from http://www.ielts.org/PDF/Guide_Edu-%20Inst_Gov_2013.pdf

Knoch, U., Rouhshad, A., Oon, S.P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing, 28*, 39-52. doi:10.1016/j.jslw.2015.02.005

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*(863). doi:10.3389/fpsyg.2013.00863

Lakens, D. (2017, January 20). Equivalence Tests: A Practical Primer for t-Tests, Correlations, and Meta-Analyses. Retrieved from osf.io/preprints/psyarxiv/97gpc

Lee, I. (2013). Research into practice: Written corrective feedback. *Language Teaching, 46*(1), 108-119. doi:10.1017/S0261444812000390

Scher, L., Kisker, E., & Dynarski, M. (2015). *Designing and Conducting Strong Quasi-Experiments in Education. Version 2*. Retrieved from https://eric.ed.gov/?id=ED561293

Shadish, W. R., & Cook, T. D. (2009). The Renaissance of Field Experimentation in Evaluating Interventions. *Annual Review of Psychology*, *60*(1), 607-629. doi:10.1146/annurev.psych.60.110707.163544

Spada, N., & Lightbown, P. M. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal*, *83*(1), 1-22.

Meihami, B., & Meihami, H. (2013). Correct I or I don't correct myself: Corrective feedback on EFL students writing. *International Letters of Social and Humanistic Sciences*, 7, 86-95. doi: 10.18052/www.scipress.com/ILSHS.7.86

Müller, A. (2011). Addressing the English language needs of international nursing students. *Journal of Academic Language & Learning*, 5(2), A14-A22.

Müller, A., Arbon, P., & Gregoric, C. (2015). A school-based approach to developing the English proficiency of EAL university students. *Journal of Academic Language and Learning*, *9*(2), A63-A78.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*(4), 375-389. doi: http://dx.doi.org/10.1016/j.jslw.2012.09.004

Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System, 37*(4), 556-569. doi: http://dx.doi.org/10.1016/j.system.2009.09.002

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23*(1), 103-110.

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*(4), 103–131. doi: 10.1111/lang.12029

Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies, 10*(2), 29-46,187. doi: 10.6018/ijes.10.2.119181

Trochim, W. M. K. (2006). *Research methods knowledge base*. Retrieved from http://www.socialresearchmethods.net/kb/index.php

Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing, 8*(2), 111-122.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16*(4), 255-272. doi: 10.1016/j.jslw.2007.06.003

Truscott, J., & Hsu, A. Y. (2008). Error correction, revision, and learning. *Journal of Second Language Writing, 17*(4), 292-305. doi:10.1016/j.jslw.2008.05.003

Weber, E. (2009). Quantifying Student Learning: How to Analyze Assessment Data. *The Bulletin of the Ecological Society of America, 90*(4), 501-511. doi:10.1890/0012-9623-90.4.501

Winchester-Seeto, T., Homewood, J., Thogersen, J., Jacenyik-Trawoger, C., Manathunga, C., Reid, A., & Holbrook, A. (2014) Doctoral supervision in a cross-cultural context: Issues affecting supervisors and candidates, *Higher Education Research & Development, 33*(3), 610-626. doi: 10.1080/07294360.2013.841648

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage.