

Learning from a corpus of students' academic writing*

Aprendiendo de un corpus de textos académicos de alumnos

Renia Lopez-Ozieblo

THE HONG KONG POLYTECHNIC UNIVERSITY
HONG KONG

renia.lopez@polyu.edu.hk

Received: 24-IX-2019 / **Accepted:** 14-XII-2020

DOI: 10.4067/S0718-09342021000200645

Abstract

This paper summarizes the results of a study based on students' academic writings. The study is grounded on Systemic Functional Linguistics (SFL) theory and pedagogy, in particular the 3x3 matrix, a SFL framework successfully implemented in Australian schools (Humphrey, Martin, Dreyfus & Mahboob, 2010) and the Teaching and Learning Cycle (Callaghan & Rothery, 1988). The study took place at the Hong Kong Polytechnic University, with 33 students from a Humanities content subject. This is an English medium of instruction higher education institution where, for most students, English is not their mother tongue. Samples of students' academic writing were analysed in the classroom to identify successful moves to express field, manner and tenor in the genre of argumentative essays. These samples were then used to provide specific writing instruction to students before they were asked to start constructing their texts independently. Feedback was provided in place of joint construction and finally students produced their final essays. Students' confidence levels on their writing abilities were measured, with knowledge surveys, at the beginning and end of the term, their final grades compared to those of the previous cohort and correlated with the end of term writing-ability confidence levels. Our results suggest that there was an increase in students' confidence in their writing abilities and that their grades were better than those of the previous cohort. When correlating grades with writing confidence levels we found two potential correlations, one positive for students with lower confidence levels and one negative for over-confident students.

Key Words: Writing Instruction, Second language writing, Knowledge Surveys.

Resumen

Este artículo resume los resultados de un estudio de corpus basado en trabajos académicos de alumnos, fundamentado en la teoría de la Lingüística Sistémico-Funcional (SFL), en particular en la matriz 3x3 (Humphrey, Martin, Dreyfus & Mahboob, 2010) y en el Ciclo de Enseñanza y Aprendizaje (Callaghan & Rothery, 1988). El estudio tuvo lugar en la Universidad Politécnica de Hong Kong, con 33 alumnos de una asignatura de Humanidades impartida en inglés, la segunda lengua de la mayoría de los alumnos. Comenzamos identificando recursos para expresar el campo, el modo y el tenor en el género de los ensayos argumentativos. Estos se tomaron como modelo y se compartieron con los alumnos antes de que creasen sus propios textos. En lugar de realizar una creación conjunta del texto con los alumnos, los profesores-investigadores les proporcionaron retroalimentación antes de la producción de los textos finales. Los niveles de confianza de los alumnos en sus habilidades de escritura se midieron a través de encuestas de conocimiento al comienzo y al final del trimestre. Se compararon sus calificaciones finales con las de la cohorte anterior y se correlacionaron con los niveles de confianza medidos. Encontramos un aumento en la confianza de los alumnos en sus habilidades de escritura y que sus calificaciones fueron mejores que las de la cohorte anterior. Al correlacionar las calificaciones con los niveles de confianza, encontramos dos correlaciones, una positiva para los alumnos con niveles de confianza más bajos y otra negativa para los alumnos con exceso de confianza.

Palabras Clave: Formación en escritura, escribir en un segundo idioma, Encuestas de Conocimiento.

INTRODUCTION

In 2017, during the evaluation of final assignments for a Humanities subject at a Hong Kong tertiary institution, it became obvious that there were wide discrepancies in students' writing competence. Students at this institution are mostly Chinese native speakers (their first language (L1) being either Cantonese or Mandarin) but the medium of instruction is English, their second language. The subject was an elective for students from all years and faculties that required writing a final argumentative essay. Students had to write this independently without any writing instruction and yet they were partly evaluated on their language ability. It was obvious that some writing instruction (WI) ought to be provided but we had little information as to how students felt about their academic writing abilities, what writing aspects ought to be targeted first and whether it would be possible to improve students' writing in the limited time available during a content subject. Based on existing successful projects (Forey, 2014) our hypothesis was that it would be possible to improve students' academic writing by providing short bursts of targeted writing instruction to address some of the key issues students' felt less confident about.

This paper details the process and results of this study. First, the language of previous A graded papers from various disciplines was analysed to identify common key strengths. The analysis followed the 3x3 matrix, a Systemic Functional Linguistics framework developed by Humphrey et al. (2010). Next, a Knowledge Survey (KS) (Nuhfer & Knipp, 2003) was administered to students to identify their confidence

levels on their ability to produce A graded papers (based on the results of the analysis). Writing instruction in the areas with the lowest confidence levels was then developed and administered to the 2018 cohort of this subject. A number of other measures were put in place to provide WI during the content lectures, including the revision of the evaluation criteria, rubrics and evaluative assignments to ensure that students were guided in the production of their final essay. For this purpose, a L2 WI pedagogy was designed based on Systemic Functional Linguistics (Halliday, 1993), following a Teaching and Learning Cycle (TLC) (Callaghan & Rothery, 1988) that focused on the deconstruction and construction of one specific text genre, argumentative (Hirvela, 2017) –also known as expository in SFL, (Martin, 1989). As the changes required significant effort, it was considered essential to be able to measure their success not only by comparing 2017 and 2018 final assignment grades but also by tracking students' self-reported confidence levels in their writing capabilities through a post-intervention Knowledge Survey.

The study adds to the understanding of the success of Systemic Functional Linguistics (SFL) genre-based pedagogical WI in an L2 (second language) context, demonstrating that students appreciate and benefit from explicit WI. It demonstrates that WI need not be complex and can be easily integrated within a content course. The WI developed is suitable for the online teaching medium and to foster autonomous learners. The study also highlights the benefits of the Knowledge Survey as a tool which is easy to implement and provides fast and valuable data. This study took place in a Hong Kong tertiary institution. However, we believe its methodology would be easily replicable in any institution where content subjects are taught in a L2.

This paper starts with a review of the existing literature on L2 writing and Knowledge Surveys. The next section describes the study in detail and it ends with a section on the results which are discussed as they are presented. The conclusion highlights our key learnings.

1. Theoretical framework: L2 writing

Traditional L2 writing instruction used to mirror the practices used in L1 (Hyland, 2002) as the writing process in both languages was considered to be the same. Both were considered to be recursive processes that involved planning, writing and revising, and thus the results were not differentiated (Silva, 1993). Later studies have proven this was an oversimplification and that adult L2 writing, at least according to L1 evaluators, is less effective than that of L1 writers. L2 writers find it harder to set writing goals and to generate and structure the material; their transcribing is less fluent (fewer words with more but shorter T-units, fewer but longer clauses); writing contains more errors and is less effective overall, with fewer subordinate or passive clauses (Silva, 1993).

Although the evidence correlating grammar accuracy and syntax complexity with good writing is not convincing (Hyland, 2011b) many students are still taught to focus on these two points without understanding the significance of their linguistic choices on creating meaning within the specific genre they are developing. Typically these students tend to use similar rhetorical structures in both their L1 and L2 and so specific L2 writing instruction is necessary to make students aware of these differences and develop their competence in the L2 writing (Leki, 2011).

Genre based writing instruction provides students with access to sample texts to deconstruct them and understand the meaning each element is adding in the particular genre. Even this is not enough, as practice, text construction, is required to assimilate new ideas and feedback is necessary to make students aware of gaps in their knowledge (Hyland, 2011a). This teaching and learning process follows a cycle of joint text deconstruction (students and teacher together), joint construction (the text is built by the student/s and the teacher, either in joint sessions or through feedback) and finally student independent construction (Callaghan & Rothery, 1988). Form-focused instruction together with feedback has been found to be successful (Ellis, 2009; Hyland, 2011a) by raising students' consciousness of how language choices can determine meaning. Hyland and Hyland (2006) reported that feedback, a key element of the writing assessment cycle, needs to cover all elements of writing: content, organization, language and style.

Although SFL metalanguage is criticized for being too complex (Bourke, 2005), its advocates argue that metalanguage is a useful tool to talk about language (Basturkmen, Loewen & Ellis, 2002; Borg, 2015), raises student' consciousness about partially acquired knowledge (Bitchener & Storch, 2016) and allows a more in-depth understanding of how "language constructs knowledge" (Gebhard, Chen, Graham & Gunawan, 2013: 108). Pessoa (2017) argues that an SFL pedagogy helps teachers by following a clear framework that allows them to make explicit the metalinguistic features needed in successful argumentation. Within the Hong Kong context a number of studies following a similar methodology have been carried out, both in schools and higher level institutions with positive results (Dreyfus, Humphrey, Mahboob & Martin, 2015; Forey, 2014).

1.1. Argumentative writing

One of the most popular and demanding writing genres in tertiary education is that of argumentative essays (Wingate, 2012), a genre requiring further discussion within L2 writing (Hirvela, 2017). Students have been reported to perform worse in this type of essay than in narrations, reflected in shorter texts, poorly structured and where the argument is not adequately supported, partly due to a lack of understanding of genre requirements (Wingate, 2012). Byrnes (2013) points out that writing is not an innate skill and has to be developed through instruction. To write strong argumentative

essays, students need to be explicitly shown the tools, allowed to practice them and be provided with feedback (Byrnes, 2013; Hyland, 2007; Lee & Deakin, 2016).

Less experienced writers have difficulties identifying and introducing a topic, acknowledging others' ideas and stating their own position (Schleppegrell, 2013), struggling to display their own knowledge of the topic (ideational metafunction), presenting their knowledge with authoritativeness, managing the relationship between reader and writer (interpersonal metafunction), organizing the text and making it coherent (textual metafunction) (Hyland, 2002). The resources needed to develop these three metafunctions can be analysed by genre, semantic or grammar level, providing educators with a 3x3 matrix (Humphrey et al., 2010) of resources to develop writing literacy at each of the levels. Some of the key resources included in writing strong expositions (as well as other academic genres) include: nominalization to formulate more abstract and condensed ideas, develop logical arguments and evaluation (Humphrey, 2017); theme and rheme, to clarify what the topic is and what the writer says about it (Hyland, 2002) and develop explicit connections between ideas (Jenkins & Pico, 2006); adequate tenses and modality to express various degrees of probability (Jenkins & Pico, 2006); and learning how to add others' voices into their writing. In addition, students need to know how to identify key relevant resources and summarize their comments.

Based on SFL genre-based pedagogical practices described above, we developed WI material to teach students how to write an argumentative essay. The material was based on assignments written by a previous cohort. The deconstruction and construction sub-cycles of the TLC had to be amended to fit the time constraints and some of the metalanguage was avoided for the same reason. However, extensive feedback was given, to deconstruct and reconstruct students' submissions throughout the term, mostly coded following what had been covered during the instruction. Our expectation was that students' writing would improve, compared to writings from the previous cohort, and that students would be more confident in their academic writing skills. To test confidence levels we employed Knowledge Surveys.

1.2. Knowledge Surveys

Knowledge Surveys (KS) were developed by Nuhfer and Knipp (2003), based on their earlier work (Knipp, 2001; Nuhfer, 1996), as a tool for students to identify their knowledge confidence levels, also providing teachers with a powerful course organization tool. A KS lists the topics to be covered during the course and students indicate their perceived level of knowledge on the particular topic. The KS is usually performed at the beginning and the end of a course, which has the dual benefit of allowing students to revise the content they should know and for teachers to identify potentially weaker areas of knowledge. It is considered a more efficient tool than traditional evaluations (Bowers, Brandon & Hill, 2005), in that it can be completed

relatively fast, as students are not writing the answers but rating their confidence level on their ability to answer the question. Most KS deploy a three point rating scale (1: I don't know how to do this; 2: I could do this with help; 3: I am confident I can do this) (Bowers et al., 2005; Nuhfer & Knipp, 2003), although other authors have used four or five point scales (Luce & Kirnan, 2016; Favazzo, Willford & Watson, 2014).

KS can be helpful tools to measure the effectiveness of new instruction tools or pedagogies (Wirth & Perkins, 2005). However, their adequacy as a tool to predict academic performance has been questioned. Some studies report increased levels of confidence that were variably matched with other evaluation tools such as exams or assignments (Bowers et al., 2005; Luce & Kirnan, 2016). In other cases, not only did confidence increase but there was a good correlation with the final grades (Wirth & Perkins, 2005; Favazzo et al., 2014), although none of these authors advocates using KS as the sole evaluation tool for courses. Trying to explain the weak correlations between levels of confidence and grades, Bowers et al. (2005) speculated that students might have difficulties evaluating their own confidence levels for topics that go beyond Bloom's (1956) initial levels of recall, comprehension and application. This hypothesis was tested by Clauss and Geedey (2012) who reported that students were better at self-evaluating their confidence levels for the lower and higher Bloom levels.

Another factor to take into account was reported by Kruger and Dunning (1999) who studied each individual's ability to recognize their own skills and knowledge. Their study observed that low performers tended to overestimate their performance, while high performers tended to underestimate it. These observations have been corroborated by other studies where overconfident students scored lower grades (Luce & Kirnan, 2016; Bell & Volckmann, 2011).

2. Methodological framework

The objective of this study was to ascertain whether explicit instruction in English academic writing would translate into an enhancement of the writing performance of students, and of their writing confidence levels, in a content class in an English language tertiary education institution in Hong Kong, where English is an L2. Based on the existing literature on L2 writing (Byrnes, 2013; Hyland, 2007; Lee & Deakin, 2016), our hypothesis was that explicit WI would translate into better-written essays.

The need for the study became obvious after realizing the difficulties that students of a particular Humanities subject (a non-proficiency subject) experienced when confronted with the end-of-semester assignment. This was very obvious in the end of term results of the 2017 cohort 2500-word essays on socio-cultural topics selected by the students. Aside from plagiarism issues students had difficulties selecting appropriate references, integrating voices from other texts into their own, evaluating others and their own ideas, correctly referencing them and, overall, in writing in a coherent and cohesive manner. This particular subject was an elective open to all

faculties and years and it is part of the writing requirements for all students. As it was a culture related content subject, the syllabus did not contain any elements of learning how to write. Students were expected to know how to write academic papers and avoid plagiarism by the time they took this subject. However, reality suggested otherwise.

2.1. Participants

The students enrolled in the above mentioned subject were asked to participate in this study. The number of students in 2017 had been 59 but in 2018 was 33, making it feasible to conduct a study that would involve extensive feedback to students. Two content teachers (one, the PI of this project) and two additional project officers (both with a focus on academic writing and one also a teacher at the institution) were involved in the design and evaluation of the rubrics, writing content and assignments. An expert on rubrics on language writing was asked to validate the rubrics created and an additional independent evaluator was also employed at the end of the term.

During the third week of term students were notified of the project and asked to provide their consent to participate, this included giving access to evaluate their work for research purposes as well as recording parts of the class. Two students did not provide their consent and their material has not been included in this study.

2.2. Process

The main writing-related issues to be addressed were identified after analyzing a corpus of assignments from a previous cohort, conducting a series of interviews with seven teachers in three different departments within the institution, three interviews with students and observing a workshop on academic writing. These issues were: incomprehensible text with an unclear message due to overly long, convoluted sentences; lexico-grammatical problems, including errors in tenses, repetition of informal oral English; lack of logic, impacting coherence and cohesion; and inadequate selection and referencing of sources.

Based on these issues, writing instruction (WI) content was developed to address a number of specific points that were considered to be key to start addressing the difficulty. The WI content was grouped into six 10 minute presentations that were delivered during the scheduled lecture time –two hours long– usually just before the break. In addition to the presentation, students were asked to go online and attempt four to five writing related exercises based on what they had just heard. Students did not have to complete the exercises but all of those attending the lectures did.

For the 2018 cohort, the evaluation of the subject was slightly modified from that of the previous year (within the existing constraints of the institution) to work with the students throughout the semester and help them in the deconstruction of texts

and construction of their writing. This meant that instead of a final 2500 word assignment on a new topic, students had to write four preparatory short assignments before writing the final one (1200 words), all on the same topic. To evaluate the impact of our WI content, at the beginning and at the end of the term students were asked to fill in a Knowledge Survey (Nuhfer & Knipp, 2003) to identify changes in their level of confidence about the various writing issues.

The WI content was designed to anticipate difficulties that students could encounter when carrying out the preparatory assignments. To prepare for the final assignment, a comparison of a socio-cultural topic from two different countries' points of view, students first had to:

- Provide three potential socio-cultural topics they would like to explore, with a short summary for each.
- Choose one of the topics and a country currently experiencing/dealing with it and find three relevant sources of information, and summarize the key ideas in those sources.
- Find another three relevant sources with different/opposing ideas to the previous ones –including academic ones– and again summarize the key ideas.
- Taking into account all the sources examined, provide an evaluation of the issue that included the various authors' voices as well as the student's own voice.

Finally, students had to select another country experiencing/dealing with the issue, find relevant sources, evaluate these and compare the issue in the two countries –this was the final assignment. For further detail on each assignment and evaluation criteria, please go to: <https://englishpolyu.wixsite.com/renialopez/academic-writing>.

The WI topics were matched as follows:

Assignment	WI topic
Select three topics	Register
Summaries of sources 1	Nominalization
Summaries of sources 2	Summarizing and paraphrasing
Evaluation	Cohesion
Final assignment, comparison	Citations and Evaluation
	Comparison and key points

The WI content followed the 3x3 matrix (Humphrey et al., 2010) and was designed to minimize the need for excessive metalanguage. This decision was taken after realizing at the beginning of term that only one student out of 33 was confident as to what 'Academic English' meant (similar observations are mentioned in Borg (2015)). The theoretical framework behind this teaching practice was the Teaching and Learning Cycle (Callaghan & Rothery, 1988). As the subject was not a writing

proficiency one, for practical reasons it was not possible to carry out a full deconstruction and reconstruction in class with the students. We compromised by initiating the deconstruction process for the students (with the WI content given) and then encouraged further independent deconstruction (with the various online exercises). Students then started the construction process by themselves by writing the fortnightly assignment and teachers participated in the construction process by providing extensive feedback. Linguistic and content feedback was given in written format mostly, through the learning management system of the institution. In some cases (12% out of 33), students asked for face to face detailed feedback as well.

The study evaluated the final assignment produced by these students and compared the overall grades to those obtained the previous year by students of the same subject who did not have any structured writing guidelines. This was done with an independent t-test and bootstrapping measures. The study also compared the results from the pre-semester and post-semester knowledge surveys, through a paired-t test and finally it explored the idea of the post-semester KS being a potential tool to predict assignment grades, through a robust correlation. All statistics were calculated using R (a statistics software). Cohen's d for effect size was calculated using the tools found at <http://www.cognitiveflexibility.org/effectsize/>. The reporting of the statistical tests used includes detailed descriptions of the assumptions and normality checks carried out, as well as visual information, as advocated by Larson-Hall (2015).

2.3. Knowledge survey

The knowledge survey was designed to make students reflect on their knowledge of various issues related to writing an academic paper. Although KS tend to be quite long, to cover the extent of the content to be delivered under the subject, in this case we compromised on the length, just 29 questions, to maximize survey completion. The questions were divided into five sections and were based on the issues identified through the interviews (as mentioned above) and the 3x3 matrix (refer to <https://englishpolyu.wixsite.com/renialopez/academic-writing> for a copy of the questionnaire). The five sections were:

1. Demonstrating knowledge of the topic you are writing about (questions 1 to 6).
2. Language use (questions 7 to 13).
3. Structuring the text (questions 14 to 20).
4. Evaluating knowledge and your research (questions 21 to 27).
5. Understanding the assessment criteria (questions 28 and 29).

The first KS was carried out during week 3 of term, just before the first WI content presentation and the last one during the last week of class (five days before the final assignment was due). The KS were bilingual (Cantonese/English), to ensure students understood all items evaluated. For each question, students had to select their

confidence level. The options were three, following those presented in existing KS studies (Nuhfer, 1996; Nuhfer & Knipp, 2003; Wirth & Perkins, 2005):

1. Low confidence level: I don't know how to do this.
2. Medium confidence level: I have a rough idea but I would need to check some resources or ask around to do this.
3. High confidence level: I know how to do this without much difficulty.

The average of the responses to all questions was obtained for each student and the Pre-KS responses were compared to the Post-KS responses.

2.4. Evaluation

Students assignments were evaluated according to the rubrics (see <https://englishpolyu.wixsite.com/renialopez/academic-writing>) and given a percentage grade (based on the institution conversion tables). For this study we only took into account the grades obtained for the final assignment, in order to be able to compare the grades to those from the 2017 group. All assignments were first evaluated on the writing by the project officers, then the PI re-evaluated the writing and the content relating to the subject, giving a final grade, achieving evaluators agreement on grades in 91% of cases (the remaining 8% was discussed with the second teacher for the subject). An independent, external evaluator, assessed 20% of randomly selected assignments, in all cases giving the same or higher grades (5% higher on average). We believe the internal evaluation team might have been grading lower as they were taking into account students' responses to previous feedback (if the feedback had not been addressed). For the analysis the lowest grades, as given by the internal team, were used. The 2017 final assignments had been evaluated by the PI and the second teacher for this subject (not involved in the research), with 20% of assignments checked by both teachers obtaining 100% agreement on grades. Only the overall grades for the final assignments for the 2017 and 2018 cohorts were compared.

3. Results and discussion

This study compared the confidence levels on students' writing skills as reported through knowledge surveys (KS) before and after explicit writing instruction (WI). It compared the grades of the final assignment of these students to those obtained by the previous cohort, in 2017, and it investigated the use of the KS as a tool to predict grades.

3.1. Evaluation

The mean confidence level per student for the Pre-KS, beginning of the term, and Post-KS, end of term was calculated. The means of the two surveys were compared using a paired t-test (we had provided participants with an explanation as to how to interpret the possible answers of the Likert scale, and we believe the distance between

categories to have been equal). Twenty eight students answered the Pre-KS and twenty five the Post-KS. However, only twenty students answered both the Pre and Post-KS and consented to the data being used for research purposes.

Before carrying out the test we analyzed the distributions to check for normality and equal variances. Within each distribution the values were independent of each other, but obviously related between the Pre-KS and Post-KS. When visually inspecting the statistical summaries of the distributions and the box-plots (see Figure 1) and Q-Q plots the two sets of data seemed to be normally distributed. These distributions had similar sizes of boxes (IQR), meaning the variances were similar, the distribution of the data did not seem to be skewed and there were no outliers. To confirm normality a Shapiro-Wilk normality test was carried out, $W = 0.97$, $p\text{-value} = 0.70$ for the Pre-KS and $W = 0.96$, $p\text{-value} = 0.57$ for the Post-KS distribution, confirming a normal distribution as for both distributions $p > 0.05$ (see the Appendix for figures and other statistical data). The mean of the Post-KS ($M = 2.33$, $SD = 0.33$, $N = 20$) is higher than that of the Pre-KS ($M = 2.053$, $SD, 0.36$, $N = 20$), suggesting an improvement.

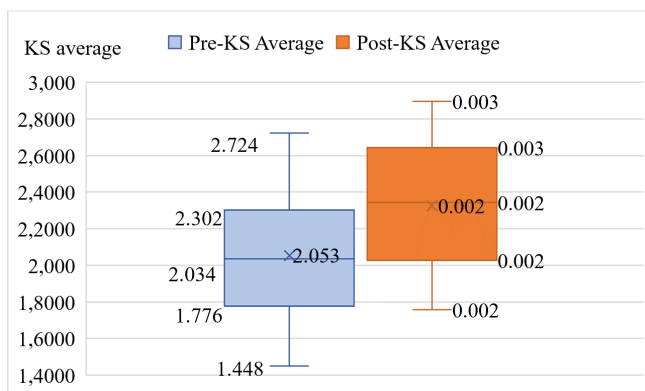


Figure 1. Pre- and Post-SK average values.

A parametric paired samples t-test found that a 95% confidence interval (CI) for the difference between the Pre-KS and the Post-KS showed a statistical and important difference [0.12, 0.42]. As this CI does not span zero, it means the difference between the two groups is statistical, and the size of the interval suggests that the actual mean difference could be as small as 0.12 or as large as 0.42. As the difference between the means could be a maximum of ± 2 , this translates to 6% to 21%, a relatively small interval, suggesting the effect was large. Therefore, we can conclude that there was an improvement in the overall confidence results and that the additional writing teaching might have made a large difference. The effect size, calculated using the pooled SDs, is large, Cohen's $d = 1.18$ showing that the measures implemented to enhance students' writing might have been very effective.

To increase the accuracy of the results, Nuhfer (2015) recommends performing a question by question comparison of KS responses, rather than taking the mean for each student. We felt it fell outside the scope of this study, as we were interested in the overall effect of all the measures implemented (rubrics, assignments, WI and feedback). However, the responses to the KS have also been analyzed by section – these have been reported in (Lopez-Ozieblo, 2021). We found an increase in confidence in four out of the five areas: ‘Demonstrating knowledge of the topic you are writing about’, ‘Language use’, ‘Evaluating knowledge and your research’ and ‘Understanding the assessment criteria’. Students still seemed to struggle with ‘Structuring the text’, as reported above.

3.2. Assignment Grades: 2018 versus 2017 results

In 2018, the final assignments were evaluated following the rubrics available in <https://englishpolyu.wixsite.com/renialopez/academic-writing>. The 2017 assignments had been evaluated following less detailed rubrics, although covering the same evaluating criteria. In 2017 there had been 61 students but only 59 final assignments were graded (the others there were found to contain plagiarized extracts and were not graded). In 2018 there were 29 final assignments submitted.

To compare these two distributions we treated them as independent, as the students were completely different, and planned to carry out an independent t-test to compare the two means. The assumptions needed for calculating a t-test are that: the dependent variable (the percentage grades with equal intervals between each grade point) should be measured in interval-level measurements, the data should be independent, normally distributed and groups should have equal variances. If this latter assumption is not met, it would be necessary to use either the Welch procedure or robust methods and should the distribution not be normal then robust procedures with means trimming or bootstrapping would need to be employed.

A check of the mean scores for 2017 ($M = 70.4$, $SD = 13.86$, $IQR = 20$) showed that they were lower than those for 2018 ($M = 81.9$, $SD = 9.44$, $IQR = 12$). From the statistical summary of the distributions and a visual check of the box-plots (see Figure 2) and QQ plots (see Appendix) it would seem that both distributions are negatively skewed, although the effect is more marked in the 2018 distribution. The variances do not seem to be equal, with the data from 2018 distributed closer to the median than that from 2017. A Shapiro-Wilk normality test indicated that the 2017 distribution was not normal (2017 $W = 0.95$ $p = 0.02$). However, the 2018 was ($W = 0.96$ and $p = 0.44$, $p > 0.05$ indicating a normal distribution).

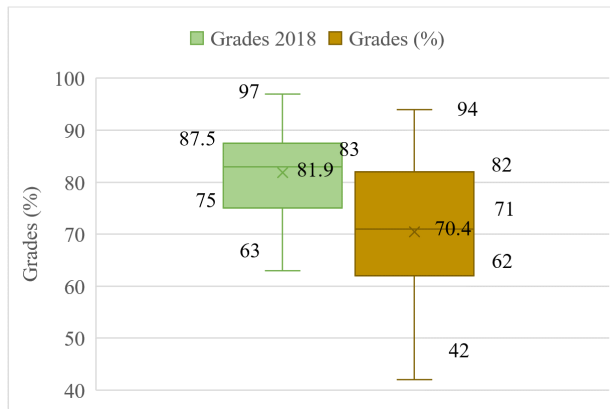


Figure 2. Grades (percentages) 2018 vs. 2017.

The variances do not seem to have homogeneity as the width of the box plots is quite different. This is confirmed numerically as the squares of the SDs is quite different as well (2017 data SD = 13.86, SD² = 192 and 2018 data SD = 9.44 SD² = 89.11). Although on testing the variances, using Levene's test for homogeneity, they were found not to be significantly different as $p = 0.074$ ($p > 0.05$ suggests no significant difference).

Following the recommendations of Larson-Hall (2015) to use robust statistics, in this case, as one of the distributions did not comply with the normality assumption, we carried out a Welch two sample t-test (the non-parametric equivalent of an independent t-test) and found strong evidence of a difference between the scores of the two means ($t = -4.54$, $df = 77.00$ $p = 0.00002$). The CI indicates that the actual difference in grades will lie, with 95% confidence between the interval (-16.45, -6.43). As this does not span zero we can reject the H₀, that there is no difference between the two data sets. This is a fairly narrow interval, 10 points over a possible 100. The effect size for this comparison, calculated with the SD of the pooled distributions can be considered high, Cohen's $d = 1.039$.

As we had found that the 2018 might not be normally distributed we also repeated the test applying bootstrapping methods. The previous results were confirmed, as the CI was similar (6.52, 16.39). Therefore, we concluded that there was a difference in grades between the two cohorts.

3.3. Lessons learnt

We believe the difference observed in the grades is in part due to the changes implemented. Among the 2017 cohort we had observed most of the issues mentioned in the literature review: students' struggled to identify topics (Schlepppegrell, 2013); texts were often just rephrases of information from other authors (Hirvela, 2017), sometimes not sourcing them at all, and it was difficult to differentiate between

students' own and others' voices (Pessoa, 2017). These aspects were particularly improved in the 2018 cohort. By guiding students in the selection of the topic and by ensuring that students persevered with that topic throughout the term, they gained a thorough understanding of the subject. This meant they were more capable in presenting different viewpoints and were more confident in expressing these as well as their own perspectives with authority (an issue identified by Lee & Deakin (2016) in Chinese ESP writers). By knowing their topics better they were also able to write about them without paraphrasing.

The pedagogical changes implemented had followed an adapted TLC (Callaghan & Rothery, 1988) where examples of the genre had been presented to students, both during the short presentations and the exercises students had to carry out. In most cases, the examples had been taken from texts from undergraduate students from this and other subjects for which we had consent. We believe that being able to use as examples real texts from students, both 'good' and 'bad', was useful as students recognized their own errors in them. The feedback provided after each assignment was extremely detailed, in most instances it was coded (not providing a correction but identifying the function of the ideal text) (Ellis, 2009) and referring to content which had been delivered in previous lectures. As the assignments progressed the detailed feedback on older topics diminished to give salience to new-topic feedback. One common issue with feedback is that students might not read it (Lee, Mak & Burns, 2015). To avoid this possibility feedback was given in such a way as to guide students' next assignment. In addition, repeated errors (ignoring feedback) were penalized and addressing feedback was rewarded. As all the feedback online was provided through the Learning Management System, it was possible to identify that all students had opened their corrected term assignments designed to guide in the writing of the final one –the corrections of final assignments were consulted by only 15 students (51%). By providing feedback to a series of guiding assignments (in practice, deconstructing and constructing students' texts) we ensured that students referred to the feedback when writing, thus enhancing students' awareness of the form and function of the language (Hyland, 2002).

There were some clear improvements: in most cases, students had initially struggled with the notion of topic sentences to introduce the main theme of the text/paragraph. However, through nominalization (which was widely used) we observed a marked improvement in the presentation of key ideas (Humphrey, 2017). There was a clear reduction of informal English overall. Sentences became shorter and clearer. The majority of students showed a clear improvement in the use of sources and how to reference them. Cases of plagiarism went down from previous years, with students seemingly aware of how to avoid it, even if just by minimally paraphrasing the sources and providing the reference. Other aspects of writing did not show significant improvements, in particular structure. However, those mentioned above

are the ones students had more chances to practice on, the first assignments were relatively short and issues with structuring were not so obvious.

Despite the results, it is essential to take into account that there are a number of external factors that might have also had an effect. As students came from all faculties there were a number of Humanities students, some enrolled in the English department, and the value of the WI to these students should have been significantly lower than to other students. In addition, some students might have been taking English language proficiency subjects under the English Language Centre and might have received additional WI. Although the KS asked respondents to refer in their answers to the writing contents learnt under the study, it is very likely that other WI received would have affected their writing confidence. This might have been reflected in the KS, probably minimizing the gains of those who had been instructed in writing before starting the course and perhaps increasing the values of those who received concurrent instruction.

3.4. Correlation between KS values and Grades

Having found that the grades of the 2018 cohort had improved and that the confidence level had also improved, we sought to find out whether there was a correlation between the Post-KS confidence levels and the grades of the final assignment. Twenty five students had filled in the Post-KS test, but one did not submit the final essay. We therefore carried out the correlation with only twenty four students.

After plotting the data to carry out an initial visual inspection, we realized that there seemed to be two groups (see Figure 3) with opposing behaviors. Obviously the correlation was non-linear and as there seemed to be a turning point just before the Post-KS median (Mean = 2.41, Median = 2.345, N = 24) we divided the data into two groups: Post-KS values <2.34 (Small group (S)) and Post-KS values >2.34 (Large group (L)).

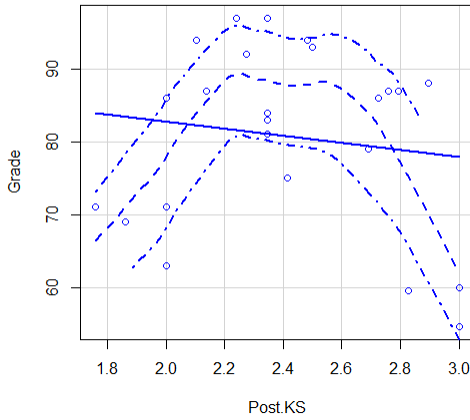


Figure 3. Scattered plot 2018 Grades vs. Post-SK values.

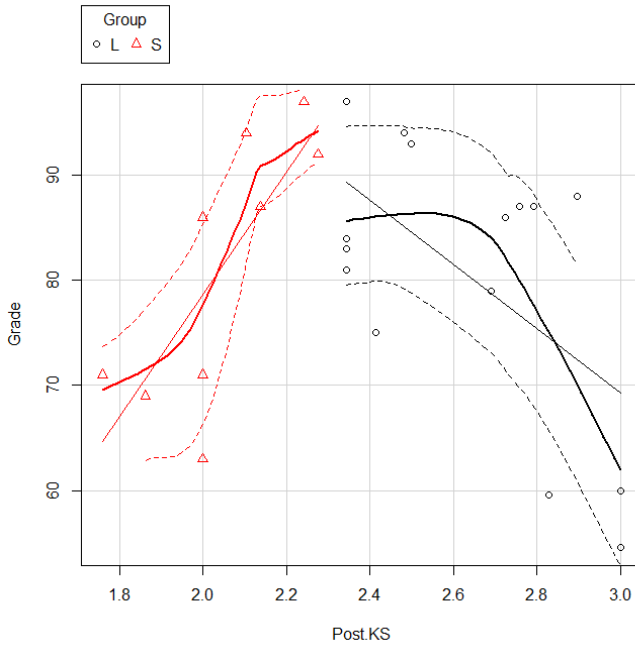


Figure 4. Scattered plot 2018 Grades vs. Post-SK values after categorising the data into Small or Large Post-KS values.

Once the data was divided into these two groups two clearer correlations (see Figure 4), one positive for lower KS values and one negative for higher KS values, were more obvious. The residuals from both groups do seem to have random values and so seem to conform to normality, although in the case of the larger Post KS

values there are a number of outliers, higher and lower values that put in question the normality (see Appendix).

A Pearson correlation (r) between the Post-KS values below 2.34 and the grades of the final assignment found the effect size of the correlation was moderate and the CI is wide (95% CI [0.01, 0.91], $r = 0.772$, $N = 9$, $R^2 = 0.596$), meaning that there seems to be a positive correlation but its coefficient might not be very reliable. The same is true of the Pearson correlation (r) between the Post-KS values above 2.34 and the grades of the final assignment. For these the effect size of the correlation was medium and the CI also wide (95% CI [-0.84, -0.08], $r = -0.58$, $N = 15$, $R^2 = 0.336$), suggesting a negative correlation but the coefficient is again not very reliable.

These results are not very strong, suggesting that the KS are not very reliable indicators of assignment grades, as pointed out by Bowers et al. (2005) and Luce and Kirnan (2016). Although we identified two groups that might be following two separate linear correlations this requires further study, ideally with larger samples. Our results seem to indicate that there might be a group of students who are under-confident (those with Post-KS values lower than 2.34) but scoring grades above 70%, and a group of over-confident students (with high Post-KS values) scoring well under 70%. Our results would corroborate those of Bell and Volckmann (2011), Kruger and Dunning (1999) and Luce and Kirnan (2016) who also found KS values and grade correlations potentially dependent on confidence attitudes. However, further information on these students would be necessary to confirm if this is indeed related to under/over confidence or if those students chose not to answer the KS truthfully. This is a possibility as even if a number of students explicitly acknowledged the efforts of the team in helping them with their writing, one also indicated that they did not consider it relevant in a content subject.

There are a number of points that need to be taken into account when assessing the validity of our results, namely that the grading of the assignments was carried out by the team involved in the study. To improve this, all assignments for both 2017 and 2018 ought to be graded by an independent evaluator. In addition, the 2017 final assignment was longer which might have affected the final grades for the cohort. Although we assumed that above a pass grade (40%) our scale was metric, with equidistant points, grading has to be considered to be somewhat subjective. The distance between the answers in the KS Likert-scale was similarly considered to be equal, this could be argued to be somewhat subjective as well, however, as an explanation was provided for what each point in the scale meant we are confident the results are valid. In addition, as the KS scale is limited to three points this might be causing ceiling and floor effects which might affect the correlation KS and grades. However, this is a limitation of the original KS (Knipp, 2001; Nuhfer, 1996).

CONCLUSIONS

The objective of this study was to assess the success of the efforts involved in preparing Systemic Functional Linguistic (SFL) based writing content (WI) that followed a Teaching and Learning Cycle (ILC). The participants of the study were tertiary students in a Hong Kong institution, writing argumentative essays in a second language (L2), in this case English, for a Humanities content subject. This quantitative study used the final assignment grades of the previous cohort, 2017, as a control sample, and implemented a series of changes in evaluation, rubrics and WI with the 2018 cohort, also based on the corpus collected from the 2017 cohort. To identify changes in the level of writing confidence of the 2018 cohort students were asked to fill in a Knowledge Survey (KS) at the beginning and end of term. A qualitative analysis of students' texts can be found in Lopez-Oziblo (2021).

Our results suggest that the changes implemented led to an improvement in writing confidence, levels and grades. Students referred to the feedback provided when writing subsequent assignments and the improvement in their writing showed a conscious effort to address the points raised in the feedback. Although the study is limited by the lack of a control group and the small sample that completed both KS we observed an improvement in the writing quality of most students. Thus we conclude that, at least for this group of students, a combination of WI and explicit feedback, together with a well-designed set of rubrics and clear evaluation criteria does help improve students writing, confirming that explicit writing instruction, within sound pedagogical practices, is beneficial to students (Byrnes, 2013; Hyland, 2011a; Schleppegrell, 2013; Luke, 2014). Some of the L2 weaknesses addressed by this study are not unique to L2 writers, even L1 university students are not always competent academic writers within their fields (Swales, 2011). We encourage all content teachers to pay attention to language as well as content. The material used for this study can be found at <https://englishpolyu.wixsite.com/renialopez/academic-writing>.

Aside from those covered above, there are, however, other factors that affect writing development, apart from pedagogy and innovative feedback methods, such as engagement (Li, 2018), motivation and students' goals (Hyland, 2011a) and the support of the institution (Byrnes, 2011). These have not been addressed in this study but that is not to say they are not important.

REFERENCES

- Basturkmen, H., Loewen, S. & Ellis, R. (2002). Metalanguage in focus on form in the communicative classroom. *Language awareness*, 11(1), 1-13.
- Bell, P. & Volckmann, D. (2011). Knowledge surveys in general chemistry: Confidence, overconfidence, and performance. *Journal of Chemical Education*, 88(11), 1469-1476.

- Bitchener, J. & Storch, N. (2016). *Written corrective feedback for L2 development*. Bristol: Biffalo.
- Bloom, B. S. (1956). *Taxonomy of educational objectives (Vol. 1). Cognitive domain*. New York: McKay.
- Borg, S. (2015). *Teacher cognition and language education: Research and practice*. London: Bloomsbury Publishing.
- Bourke, J. M. (2005). The grammar we teach. *Reflections on English language teaching*, 4, 85-97.
- Bowers, N., Brandon, M. & Hill, C. D. (2005). The use of a knowledge survey as an indicator of student learning in an introductory biology course. *Cell biology education*, 4(4), 311-322.
- Byrnes, H. (2011). Beyond writing as language learning or content learning. *Learning-to-write and writing-to-learn in an additional language*, 133-157.
- Byrnes, H. (2013). Positioning writing as meaning-making in writing research: An introduction. *Journal of Second Language Writing*, 2(22), 95-106.
- Callaghan, M. & Rothery, J. (1988). *Teaching factual writing: A genre-based approach: The report of the DSP literacy project, Metropolitan East Region*: Metropolitan East Disadvantaged Schools Program.
- Clauss, J. M. & Geedey, C. K. (2012). Knowledge surveys: Students ability to self-assess. *Journal of the Scholarship of Teaching and Learning*, 10(2), 14-24.
- Dreyfus, S. J., Humphrey, S., Mahboob, A. & Martin, J. R. (2015). *Genre pedagogy in higher education: The SLATE project*. Palgrave Macmillan.
- Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, 1(1).
- Favazzo, L., Willford, J. D. & Watson, R. M. (2014). Correlating student knowledge and confidence using a graded knowledge survey to assess student learning in a general microbiology classroom. *Journal of microbiology & biology education*, 15(2), 251.
- Forey, G. (2014). Learning non-language subjects through English: The role of language and beyond. Proceedings of the *English Enhancement Scheme and Refined English Enhancement Scheme – From Implementation to Sustainability*. Kowloon Technical School, Sham Shui Po, Hong Kong.
- Gebhard, M., Chen, I. A., Graham, H. & Gunawan, W. (2013). Teaching to mean, writing to mean: SFL, L2 literacy, and teacher education. *Journal of Second Language Writing*, 22(2), 107-124.

- Halliday, M. A. (1993). Towards a language-based theory of learning. *Linguistics and education*, 5(2), 93-116.
- Hirvela, A. (2017). Argumentation & second language writing: Are we missing the boat? *Journal of Second Language Writing*, 36, 69-74.
- Humphrey, S. (2017). *Academic Literacies in the Middle Years A Framework for Enhancing Teacher Knowledge and Student Achievement*. New York, London: Routledge.
- Humphrey, S., Martin, J. R., Dreyfus, S. & Mahboob, A. (2010). The 3x 3: Setting up a linguistic toolkit. In A. Mahboob & N. Knight (Eds.), *Applicable linguistics* (pp. 185-195). London: Bloomsbury.
- Hyland, K. (2002). *Teaching and researching: Writing*. Harlow, England: Longman.
- Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16(3), 148-164.
- Hyland, F. (2011a). The language learning potential of form-focused feedback on writing. Students' and teachers' perceptions. In R. Manchón (Ed.), *Learning-to-Write and Writing-to-Learn in an Additional Language* (pp. 159-179). Amsterdam/Philadelphia: John Benjamins.
- Hyland, K. (2011b). Learning to write: Issues in theory, research and pedagogy. *Learning-to-write and writing-to-learn in an additional language*, 17-36.
- Hyland, K. & Hyland, F. (2006). *Feedback in second language writing: Contexts and issues*. Cambridge: Cambridge University Press.
- Jenkins, H. H. & Pico, M. L. (2006). *SFL and argumentative essays in ESOL*. Proceedings of the 33 International Systemic Functional Congress, Sao Paulo, Brazil.
- Knipp, D. (2001). *Knowledge surveys: What do students bring to and take from a class?* United States Air Force Academy Educator: Spring.
- Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6), 1121.
- Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R*. Routledge.
- Lee, I., Mak, P. & Burns, A. (2015). Bringing innovation to conventional feedback approaches in EFL secondary writing classrooms: A Hong Kong case study. *English Teaching: Practice & Critique*, 14(2), 140-163.

- Lee, J. J. & Deakin, L. (2016). Interactions in L1 and L2 Undergraduate Student Writing: Interactional Metadiscourse in Successful and Less-Successful Argumentative Essays. *Journal of Second Language Writing*, 33(C), 21-34.
- Leki, I. (2011). Learning to write in a second language multilingual graduates and undergraduates. *Learning-to-write and Writing-to-learn in an Additional Language expanding genre repertoires*, 85-109.
- Li, W. (2018). *Translanguaging as a practical theory of language: Implications for language learning and research*. Proceedings at the Faculty of Humanities Distinguished Lecture Series. The Hong Kong Polytechnic University, Hung Hom, Hong Kong.
- Lopez-Ozieblo, R. (Jan 2021). Improving second language writing across the disciplines: Resources for content teachers. In M. Carrió-Pastor & B. Bellés-Fortuño (Ed.), *Teaching Language and Content in Multicultural and Multilingual Classrooms. CLIL and EMI Approaches* (pp. 191-222). London: Palgrave MacMillan.
- Luce, C. & Kirnan, J. P. (2016). Using indirect vs. direct measures in the summative assessment of student learning in Higher Education. *Journal of the Scholarship of Teaching and Learning*, 16(4), 75-91.
- Luke, A. (2014). On explicit and direct instruction. *Australian Literacy Association Hot Topics*, 1-4.
- Martin, J. R. (1989). *Factual writing: Exploring and challenging social reality*. Oxford: Oxford University Press.
- Nuhfer, E. B. (1996). The place of formative evaluations in assessment and ways to reap their benefits. *Journal of Geoscience Education*, 44(4), 385-394.
- Nuhfer, E. B. (2015). Clarification to Points in Correlating Student Knowledge and Confidence Using a Graded Knowledge Survey to Assess Student Learning in a General Microbiology Classroom. *Journal of microbiology & biology education*, 16(2), 125.
- Nuhfer, E. B. & Knipp, D. (2003). The Knowledge Survey: A tool for all reasons. *To improve the academy*, 21(1), 59-78.
- Pessoa, S. (2017). How SFL and explicit language instruction can enhance the teaching of argumentation in the disciplines. *Journal of Second Language Writing*, 36, 77-78.
- Schleppegrell, M. J. (2013). The role of metalanguage in supporting academic language development. *Language Learning*, 63(s1), 153-170.

- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *Tesol Quarterly*, 27(4), 657-677.
- Swales, J. M. (2011). Coda: Reflections on the future of genre and L2 writing. *Journal of Second Language Writing*, 20(1), 83-85.
- Wingate, U. (2012). 'Argument!' helping students understand what essay writing is about. *Journal of English for Academic Purposes*, 11(2), 145-154.
- Wirth, K. R. & Perkins, D. (2005). Knowledge surveys: An indispensable course design and assessment tool. Proceedings of the *Innovations in the Scholarship of Teaching and Learning*. St. Olaf College, Northfield, Minnesota, USA.

*ACKNOWLEDGEMENTS

This study is based on the proposal won by Dr Gail Forey and funded by the Teaching and Learning Fund of the Hong Kong University Grants Committee (2017). It was supported by the Research Centre for Professional Communication in English (RCPCE) of the Department of English of the Hong Kong Polytechnic University.

I am grateful to all the participants who have made this study possible, in particular to students who gave us access to their writing and the colleagues who agreed to be interviewed. I am also grateful to colleagues from the ELC, Dr Julia Chen and Dr Grace Lim and Dr Josephine Csete from EDC and to the members of the project team: Dr Eric Cheung, Mary Johannes and Cyril Lim.

APPENDIX – FIGURES AND STATISTICAL CALCULATIONS

Please refer to the main text for the interpretations of the results below.

Changes in Confidence Level Pre-KS vs Post-KS

	<i>Mean</i>	<i>SD</i>	<i>IQR</i>	<i>N</i>	Shapiro Wilk normality test	
					<i>W</i>	<i>p</i>
Pre KS	2.053	0.36	0.47	20	0.96764	0.7045
Post KS	2.33	0.33	0.47	20	0.96125	0.5691

Paired t-test $t = 3.7439$, $df = 19$, $p = 0.001375$

Mean of the differences = 0.27; 95% CI = [0.12, 0.43]; Pooled SDs

Cohen's $d = 1.183$

Assignment grades: 2018 versus 2017 results

Grades for	Mean	SD	SE (mean)	IQR	Skewness	Kurtosis	N
2017	70.4	13.86	1.80	20	-0.41	-0.65	59
2018	81.9	9.44	1.75	12	-0.21	-0.78	29

Shapiro-Wilk normality test

Cohort	W	p	p adjusted (Holm method:
2017	0.95234	0.02165	0.043307
2018	0.96525	0.4391	0.439131

Welch Two Sample t-test - H0: true difference in means is equal to 0

t = -4.5455; df = 77.008; p = 0.00002; 95 % CI = [-16.449 -6.427]

Reject H0: true difference in means is not equal to 0

Correlation between KS values and Grades

KS	Mean	Median	N	Pearson r	*Effect size R ²	bias	SE	Bootstrapped CI
<2.34	2.41	2.345	9	0.772	0.596	0.012	0.128	[0.01, 0.91]
>2.34	80.75	85.00	15	-0.580	0.336	0.047	0.208	[-0.84, -0.08]

* small effect $r < 0.3$, medium effect $0.3 < r < 0.5$, moderate effect $0.5 < r < 0.7$, large effect $r > 0.7$