
The relationship between lexical diversity and EFL writing proficiency

XUAN WANG

University of Sydney

ABSTRACT

This study investigates the lexical dimension of EFL writing by exploring the relationship between lexical diversity and EFL writing proficiency. The relationship is explored through an analysis of forty-five email texts written by Chinese high school students that were taken from a Chinese National Matriculation English Test (NMET) practice writing test. Type-token ratio (TTR) and D (a measurement of lexical diversity through curve fitting) were used to establish indices of lexical diversity. The results showed that the lexical diversity of email texts graded at higher proficiency levels was not significantly different from the lexical diversity of email texts graded at lower proficiency levels, and that neither the TTR nor D measures had a statistically significant relationship with participants' writing scores. These findings suggest that lexical diversity measures do not always offer a reliable basis for distinguishing between writing proficiency levels.

INTRODUCTION

Vocabulary is an essential component of second language learning, showing important correlations with both grammar and skills such as listening, reading and writing (Alderson, 2005). Vocabulary plays a significant role in the complex process of writing, which involves the coordination of many higher level and lower level cognitive skills (Hayes, 1996; Olinghouse & Leaird, 2009; Scardamalia & Bereiter,

Address for correspondence: Xuan Wang, Faculty of Education and Social Work, University of Sydney, NSW 2006, Australia; Email: xwan9056@uni.sydney.edu.au

University of Sydney Papers in TESOL, 9, 65-88.

©2014 ISSN: 1834-3198 (Print) & 1834-4712 (Online)

1987). It is particularly important for second and foreign language writing. Schoonen, van Gelderen, Stoel, Hulstijn, and de Glopper (2011) found that vocabulary knowledge correlated more highly (.63) with EFL writing proficiency than with L1 writing proficiency (.47). The current study focuses on one aspect of vocabulary, lexical diversity, as lexical diversity is a commonly used scoring criterion for English proficiency tests such as IELTS and TOEFL. The relationship between lexical diversity and EFL writing proficiency levels has only been explored for a limited range of genres. The current study investigates the relationship between lexical diversity and writing proficiency in the genre of email texts. The study aims to provide insights into whether measurements of lexical diversity provide an accurate and efficient indication of the writing proficiency of EFL learners in this genre.

LITERATURE REVIEW

Lexical diversity

One of the current concerns within the literature on lexical diversity is providing an adequate definition, and specifically distinguishing between lexical diversity and lexical richness. Some researchers consider lexical richness to be synonymous with lexical diversity (e.g., Arnaud, 1984; Wimmer & Altmann, 1999), while others define lexical diversity and lexical richness as two distinct concepts with different focuses (e.g., Laufer, 2003; Read, 2000). In this study, lexical richness and lexical diversity are treated as two separate concepts. Lexical richness is viewed as referring to the quality of lexical knowledge and includes both depth and breadth of knowledge (Read, 2000). Lexical diversity is viewed as referring to the range of vocabulary in a piece of writing and includes only the breadth of vocabulary knowledge (Malvern, Richards, Chipere & Duran, 2004).

Quantifying the level of lexical diversity involves identifying how often different words are used in a text. This requires comparison between types and tokens in the text. Tokens describe the total number of words in a text, while types represent the number of different items (Nation, 2001). As a measure of the range of vocabulary used by a writer, lexical diversity has traditionally been

calculated using type-token ratio (TTR) (Malvern & Richards, 2002), which describes the ratio between the number of different words (types) and the total number of words (tokens). For instance, there are 14 tokens but 12 types in the following sentence: *I have been learning English for nine years so I can speak English fluently*. The two occurrences of *I* and *English* each belong to the same type. The type-token ratio is calculated and reported as a percentage using the following formula:

$$\text{type-token ratio} = \frac{\text{number of different words (types)}}{\text{total number of words in text (tokens)}} \times 100$$

However, a significant weakness of the TTR method when it is used to compare texts is the sensitivity of the measure to variability in text length (Nation & Webb, 2011). As a text gets longer, there are fewer chances for new types to appear, because a high proportion of the frequent types have already appeared. Thus, longer texts tend to have lower type-token ratios as an artefact of text length.

Duran, Malvern, Richards and Chipere (2004) addressed this weakness of the TTR by proposing an alternative measure of lexical diversity, *D*, which involves drawing multiple word samples from a text and plotting their respective TTRs on a curve. This process allows the relative lexical diversity of even quite short texts to be determined. Duran *et al.* (2004) suggested that if a text has a higher *D*, its lexical diversity tends to be greater. *D_Tools* (Meara & Miralpeix, 2008) is a computer program that is freely available on the *lognostics* website (<http://www.lognostics.co.uk>). It provides a standardized procedure for measuring *D* in input text files (.txt). *D_Tools* generates 100 samples of 35 randomly selected words from a text, and calculates a type-token ratio for each of these. These 100 values are then averaged to produce a composite mean ratio for all 100 samples. The program goes on to do the same calculation for samples of 36 randomly-selected words, then for 37, then for 38, and so on, all the way up to samples of 50 words. The end result is a list of 16 TTR ratios for the 35–50 word samples. These ratios form a curve,

which is then compared to the number of theoretical curves generated by the D formula. The formula for these calculations is as follows:

$$\text{TTR} = \frac{D}{N} \left[\left(1 + 2 \frac{N}{D} \right)^{\frac{1}{2}} - 1 \right]$$

Because *D_Tools* carries out its curve-fitting procedure on a curve segment of 35–50 tokens, a minimum of 50 valid words is needed to supply all 16 data points. D can be used to indicate the lexical diversity of samples with more than 50 tokens, and it does so most effectively for texts whose tokens range from 100 to 400 (McCarthy & Jarvis, 2007). D has been found to be a better indicator of lexical diversity than other traditional measures, and it has been widely used in linguistic research in both first and second language studies (Klee, Stokes, Wong, Fletcher & Gavin, 2004; Read & Nation, 2006; Yu, 2009).

The relationship between lexical diversity and foreign language writing proficiency

Studies on lexical diversity have focused primarily on the measurement of lexical diversity, so the relationship between lexical diversity and language proficiency has received less attention. The relationship between lexical diversity and foreign language proficiency has been investigated in relation to overall language proficiency (Malvern & Richards, 2002; Nation & Webb, 2011), quality of speaking (Read, 2000; Yu, 2009) and quality of writing (Engber, 1995; Jarvis, 2002; Yu, 2009). More proficient language learners are perceived to have a larger semantic network than less proficient learners (Read, 2000). For instance, a proficient foreign language learner with a large vocabulary that can be activated and used tends to produce a greater variety of words in a language activity. In contrast, a user with only a small vocabulary and little ability to activate these words may produce fewer words in a similar language activity. Read and Nation (2006) investigated the lexical use of 88

participants with a range of oral English proficiency levels from Band 8 down to Band 4 in the IELTS Speaking Test. Their findings suggested that participants in the higher scoring bands produced a larger range of vocabulary on average than those in the lower bands in relation to both tokens and types. In terms of lexical variation, the mean D values showed a pattern of decline from Band 8 to Band 4. However, there was considerable variance within each band, particularly within Bands 7 and 6. The findings indicated that although candidates with higher oral proficiency tended on average to have more diverse vocabulary on average, lexical variation could not reliably distinguish participants by band score.

However, the relationship between lexical diversity and foreign language writing has received relatively little scholarly attention. Writing is a complex process involving generating and organizing ideas, developing writing plans, reviewing and revising what has been written, and simultaneously self-monitoring performance. A writer requires effective cognitive skills, sufficient linguistic knowledge, as well as awareness of the task environment to produce a good quality composition (Flower & Hayes, 1981; Hayes, 1996; Scardamalia & Bereiter, 1987). Although linguistic knowledge is not of fundamental importance to every stage of the above mentioned process, linguistic knowledge, especially vocabulary knowledge is one of the most important and difficult issues for EFL learners in writing (Leki & Carson, 1994). One of the main problems that EFL learners have with writing is that their vocabulary tends to be very limited, and is therefore used repetitively, limiting the extent of their expression. Consequently, there is a focus placed on increasing the vocabulary diversity in the EFL classroom. EFL learners are also actively encouraged to broaden their vocabulary knowledge, as this is often a scoring criterion in EFL language tests.

The relationship between lexical diversity and writing proficiency levels is complicated and evidence of this relationship remains inconclusive. Engber (1995) explored the relationship between lexical proficiency and the quality of ESL writing by investigating the essays of 66 ESL learners from mixed L1 backgrounds and different proficiency levels. All participants wrote on the same topic. TTR was

used to measure lexical diversity, and the measurement of lexical diversity was categorized into two types: lexical variation with error and error-free variation. In terms of lexical variation with error, Engber included all lexical items, even if the lexical item was used incorrectly. In terms of error-free variation, he deleted the lexical items with mistakes in term of both meaning and form. Engber defined lexical errors in terms of both lexical choice and lexical form. The results of his study revealed significant correlations between lexical variation and writing quality for both measures (lexical variation with errors and lexical variation without errors). However, the inclusion of writers of mixed L1 backgrounds may have influenced the test results, as the writers were not grouped in terms of language background.

Jarvis (2002) dealt with the issue of language background by grouping his test subjects according to their shared L1 backgrounds. Jarvis compared different lexical diversity measures on short narratives written by adolescent EFL learners of Finnish backgrounds, EFL learners of Swedish backgrounds and L1 writers from the US. Participants wrote a film narrative immediately after watching an 8-minute silent movie. The results indicated a clear correlation between lexical diversity and the number of years that the subjects had studied English, but the extent to which lexical diversity was associated with the quality of writing was complex. For example, the study indicated that written narratives with the highest mean D tended to have the lowest, even negative, correlations between D and writing quality. In addition, the relationship between lexical diversity was found to be influenced by language background. There were significant correlations with writing quality in the narratives of Swedish participants, but the correlations for Finnish participants and for native speakers were low and not significant.

A study by Yu (2009) provided further support for the effect of L1 background on the relationship between lexical diversity and text quality. Yu measured the relationship between lexical diversity and quality of both written and spoken tasks. Two hundred compositions on five different topics by English learners from different L1 backgrounds were analysed as written discourse and 25 composition

participants were interviewed to obtain data on spoken discourse. D was used to measure lexical diversity of both written and spoken tasks. There was a significant relationship between D and the overall quality rating of the entire sample set with mixed L1 backgrounds. However, for samples from the two largest L1 groups, Filipino and Chinese, lexical diversity was not a significant predictor of the quality of learners' written compositions. This provides a further indication that divergent results can occur from testing subjects of different language backgrounds.

Another factor that may influence the relationship between lexical diversity and writing proficiency is genre. An example of this is a study by Olinghouse and Wilson (2013), who examined the role of vocabulary in L1 writing across narrative, persuasive, and informative genres. Although their study investigates L1 as opposed to L2 writing, the study is nonetheless relevant to understanding the role of genre. The results indicated that students vary their vocabulary usage between different genres. Narrative texts had greater lexical diversity than both informative and persuasive texts. Informative texts displayed the least lexical diversity of the three types. The statistical analysis of their data indicated that the contribution of vocabulary to writing quality differed between genres. For narrative texts, lexical diversity was a unique predictor, explaining 8.4% of the variance in writing quality, while for persuasive texts, lexical diversity contributed little unique variance, only 1.4%. For informative texts, lexical diversity explained neither unique variance nor total variance. It seems that lexical diversity was a predictor of writing quality for narrative texts, but not for persuasive and informative texts.

The current study aims to explore the relationship between lexical diversity and EFL writing proficiency levels for the genre of email writing for EFL writers of Chinese background. To our knowledge, the genre of email writing has not previously received scholarly attention in relation to the relationship between lexical diversity and writing proficiency. The study specifically looks at writers of Chinese background, as previous studies have shown that the L1 background of a learner may influence the relationship between lexical diversity and the quality of EFL writing. In terms of methodology, few studies

have investigated lexical diversity using two measures in the same study. In the current study, both TTR and D are used to measure lexical diversity. The use of these two methods helps to ensure this study's reliability.

The study seeks to answer the following research question:

What is the relationship between lexical diversity and EFL writing proficiency levels for Chinese high school students in an email writing task?

METHODS

Participants and setting

The participants in this study were 45 full-time senior high school students who were studying at a public high school in Northern China. Students received approximately 4 hours per week of EFL instruction. The EFL instruction mainly emphasised reading comprehension, grammar and writing. In this educational context, writing an email is a common genre in both classroom teaching and testing. Little systematic vocabulary instruction was offered throughout the academic year. None of the participating students had been abroad for study. All were aged 18-19 years and were in their final year of senior high school, at the end of which they would take the National Matriculation English Test (NMET) for university entrance.

The NMET practice writing test

In China, English has been a compulsory subject for high school students for three decades, and students seeking to enter university must pass a national standardized English test called the National Matriculation Test (NMT). In order to prepare students for the NMET, high schools hold in-house NMET practice tests two or three times each year from March to May for students in their final year of high school.

The in-house NMET practice tests mimic the format of the NMET test, which has four sections: listening, the application of English knowledge, reading comprehension, and writing. Students are given 120 minutes to complete the whole test on paper. Students are

generally assigned to classrooms in groups of approximately thirty each, with one teacher supervising each group.

In the practice test utilised in the current study, the writing section contained two parts: error correction and guided email writing. Of the total 150 points possible on the test, the email writing section accounted for 25 points, or 16.7% of the total score. The proportion of the overall time allocated for the test spent on the email writing task was controlled by the participants. The guided email writing task required participants to introduce themselves in approximately 100 words to a foreign university in the form of an application for admission (see Appendix 1 for task instructions). The beginning and end of the email text were provided as part of the prompt and were not included in the final word count or analysis.

Scoring

Two raters scored scanned versions of participants' NMET practice tests via computer. Both raters were employed by the school to assess the writing tasks, and had more than five years of experience in assessing students' writing. Although the raters were English teachers at the school, neither of them taught the participants in this study. They were required to provide impartial scores according to strict evaluation criteria. The evaluation criteria, which were freely available to students and teachers, were related to task achievement, use of vocabulary, accuracy of grammar, and coherence and cohesion. In the evaluation criteria used by the school in this study, the scores of the email writing are categorized into five proficiency levels. Level 1 contains the scores from 21 to 25; Level 2 contains the scores from 16 to 20; Level 3 contains the scores from 11 to 15; Level 4 contains the scores from 6 to 10; Level 5 contains the scores from 1 to 5 (see Appendix 2 for evaluation criteria). These scores have been used consistently across the region over the past seven years.

The email texts were randomly assigned to the two raters, and no identifying information about the participants was included in the scanned versions of the texts. The raters finished their scoring of the texts before this study commenced, and their scores were retrieved

from the school's archival database. Interrater reliability for the two raters for the forty-five texts was ($N=2$) $r = .98$, $p < 0.001$.

Selection of texts

The 45 texts in the sample were compiled from scanned versions of the NMET practice writing tests in the participating school. The texts were selected according to three criteria.

Firstly, only texts from the three highest proficiency levels (Levels 1 to 3) were chosen for this study, because if participants had responded to every bullet point within the writing task, they were awarded a minimum of 11 points (the lowest score of Level 3). Moreover, there were only a limited number of texts in the two lowest levels (Level 4 and Level 5). By only selecting tests with scores of 11 and above, the researcher could be assured that the writing section of each test had been completed, regardless of the quality of the writing. The highest score was 23, so the scores of the selected texts ranged from 11 to 23.

Secondly, the handwriting needed to be clear. One text in Level 3 was excluded because of the poor quality of the handwriting.

Thirdly, the same number of texts were drawn from each level to allow for meaningful comparisons between the different proficiency levels. Fifteen samples from each level were chosen. Although the NMET practice writing test instructions required each composition to be around 100 words, the sample selection did not depend on any particular token number.

Transcribing data

The email-writing section of the NMET practice tests was stored in the school's database in scanned form, and each of the forty-five texts needed to be transcribed to electronic files (.txt) for data analysis. The samples were transcribed without any grammatical correction, but the spelling mistakes were manually changed according to the following principles. First, this study accepted one missing letter, one wrong additional letter, or one incorrectly placed letter as correct. For example, for *country*, various misspellings (e.g., *county*, *countcy* and *counnry*) were accepted. These words were

corrected and retained. Other misspelled words were counted as invalid and deleted. Contracted forms were separated (*I'm* became *I am*; *don't* was changed to *do not*, *I'd* to *I would*, *it's* to *it is*, and so on). Multi-word items were likewise separated (e.g., *self-introduction* → *self introduction*).

These principles for correcting spelling mistakes were based on the premise that the focus of lexical diversity is to explore how diverse the range of words used is, so therefore to fulfill this primary focus, the words themselves must be in correct and recognizable form. Words with more than one incorrect letter were omitted, as it was difficult to distinguish whether the lexical items with two or more incorrect letters was an issue of spelling, or of creating a non-existent word.

Calculating lexical diversity

Two computer programs for applied linguistic research were used to analyse the lexical diversity of the samples. These were:

- *Range* (Nation & Heatley, 1996): This program can summarise the tokens and types used in either an individual text sample or multiple samples. It was used to calculate the value of tokens and types in each individual sample, as well as for the value of tokens and types within each level, and for the entire forty-five samples as a totality. *Range*, however, does not directly provide the value of TTR for individual samples so the researcher needed to calculate TTR according to the TTR formula based on the value of types and tokens provided by the *Range* program.
- *D_Tools* (Meara & Miralpeix, 2008): *D_Tools* is a program based on research by Malvern and Richards (2002) that computes the value of D to reflect the lexical diversity of a text. Raw text can be input directly into *D_Tools*. Because of the random samples of word tokens, a very slightly different value of D is reported for each sample each time the program is run. In the current study, each sample was subject to 5 complete iterations of the *D_Tools* analyses, and the average

of these five values was returned as the final D value for each sample.

Data analysis

Quantitative analysis was employed in this study. Firstly, the 45 texts were divided into three groups to represent three writing proficiency levels according to the scoring criterion provided by the NEMT exam: scores ranging from 21 to 23 were in the Level 1; scores ranging from 16 to 20 were in the Level 2; scores ranging from 11 to 15 were in Level 3. Thus, as explained, there were three proficiency levels in the study and each proficiency level had 15 samples. Secondly, TTR and D were calculated. The next stage in the preparatory process was data entry. All data relating to TTR and D were entered into the Statistical Package for Social Sciences (SPSS). Then, the TTR, and D data were each analyzed using a one-way Analysis of Variance (ANOVA), a statistical test often used to investigate differences between the means of several groups. The one-way ANOVA were used to measure the differences between the three groups in terms of both TTR and D. Finally, post hoc comparisons using the Tukey HSD test were used to test whether the lexical diversity measures for higher proficiency levels were significantly different from those of lower proficiency levels.

RESULTS

Lexical output

Analysis using the *Range* program showed that the forty-five texts yielded 5651 tokens and 798 types in total. The number of tokens in each text ranged from 86 to 156. The number of types in each text ranged from 56 to 104. *Range* also evaluated how many tokens and types were produced in the texts grouped into Levels 1-3, as total and mean scores (see Table 1).

TABLE 1
Lexical output by levels (Range analysis)

Levels	Total		Mean	
	Tokens	Types	Tokens	Types
Level 1	1946	458	129.73	85.27
Level 2	1957	436	130.47	80.73
Level 3	1749	441	116.60	75.20

Analysis of TTR

The TTR values for the 45 texts ranged from 54.6% to 75.6%. Participants within the same level showed considerable variation in TTR. For instance, the sample with the highest TTR (75.6%) was Level 2 and the sample with the lowest TTR score (54.6%) was also Level 2.

The mean TTR value of participants in Level 1 (the highest proficiency level in the study) was 64.6%, ($SD = .031$); the mean TTR value of participants in the Level 2 (the middle proficiency level) was 64.3%, ($SD = .064$); the mean TTR value of participants in the Level 3 (the lowest proficiency level in the study) was 66.9%, ($SD = .025$). A one way ANOVA was used to compare TTR values of Level 1, Level 2 and Level 3. Results indicated that TTR values did not significantly differ across the three levels, $p = .218$, $F(2, 42) = 1.58$ (see Table 2).

TABLE 2
ANOVA analysis related to TTR

Source	Sum of Squares	df	Mean Square	F	Sig.
Between groups	.006	2	.003	1.582	.218
Within groups	.079	42	.002		
Total	.085	44			

Post hoc comparisons using the Tukey HSD test indicated that the mean TTR value for Level 1 was not significantly different than that for Level 2, $p = .978$, 95% *CI* [-.0352, -.0416]. Also, the mean TTR value for Level 1 did not significantly differ from that for Level 3, $p = .335$, 95% *CI* [-.0610, -.0158]. The mean TTR value for Level 2 did not significantly differ from that for Level 3, $p = .244$, 95% *CI* [-.0642, -.0126]. The findings indicated that the TTR values of texts within the higher proficiency levels were not significantly different from that of lower levels.

Analysis of D

The D values for the forty-five texts ranged from 40.958 to 104.580. The sample with the highest D value (104.580) was in the highest level (Level 1), and the sample with the lowest D (40.958) was in Level 2. Mean D value of participants in the Level 1 (the highest group) was 69.23, ($SD = 17.57$); mean D value of participants in the Level 2 (second highest group) was 67.85, ($SD = 20.41$); mean D value of participants in the Level 3 (the lowest group) was 80.01, ($SD = 15.01$). A one-way ANOVA was used to compare D values of Level 1, Level 2 and Level 3. Results indicated that D values did not significantly differ across the three levels, $p = .135$, $F(2,42) = 2.10$ (see Table 3).

TABLE 3
ANOVA analysis related to D

Source	Sum of Squares	df	Mean Square	F	Sig.
Between groups	1330.105	2	665.053	2.099	.135
Within groups	13309.083	42	316.883		
Total	14639.188	44			

Post hoc comparisons using the Tukey HSD test indicated that the mean D value for Level 1 was not significantly different than that for Level 2, $p = .975$, 95% *CI* [-14.41, 17.17]. Also, the mean D value for Level 1 did not significantly differ from that for Level 3, $p = .233$, 95% *CI* [-26.57, 5.012]. The mean D value for Level 2 did not significantly

differ from that for Level 3, $p = .160$, 95% *CI* [-27.95, 3.63]. The findings indicated that the D values of compositions within the higher proficiency levels were not significantly different from that of lower-level compositions.

DISCUSSION

This study has investigated the relationship between lexical diversity and EFL writing proficiency levels of Chinese high school students for a previously unexamined genre, email writing. The study has used two measures of lexical diversity, TTR and D, to measure the lexical diversity of email texts produced as part of an English as a foreign language practice test, and to examine whether lexical diversity in these email texts was significantly different for higher and lower proficiency levels. The study has shown that for this email writing task neither measure of lexical diversity showed significant differences between different proficiency levels. Thus, the evidence suggests that lexical diversity in this email writing task did not offer a measure for distinguishing between higher and lower proficiency levels. This finding corroborates findings from Jarvis (2002) and Yu (2009) that lexical diversity is not consistently related to text quality across genres and language backgrounds.

The lack of significant differences in lexical diversity between different proficiency levels could be explicable in terms of the genre of the writing task. The main purpose of the email writing task in this study was to apply for university entrance and participants focused on introducing themselves in approximately 100 words. Therefore, the content involved was relatively limited and quite rigidly prescribed in the task instructions. These features might have influenced participants' vocabulary use and hence the relationship between lexical diversity and writing proficiency levels.

The lack of relationship between lexical diversity and writing proficiency levels for Chinese participants supports the results of Yu's (2009) study. Further research, however, is still required to determine why this relationship does not appear to be important for Chinese participants. The reason for this particular group of Chinese participants may be related to their understanding of the marking

criteria. The texts in the sample were chosen from NMET practice writing tests. The evaluation criteria for the NMET, which were freely available to all the students and teachers, were related to task response, use of vocabulary, accuracy of grammar, and coherence and cohesion. In an attempt to achieve high scores, it is possible that participants avoided using vocabulary they were unsure about and that they focussed more on accuracy of vocabulary use. Thus, the task constraints could have had a detrimental influence upon participants' use of rich and diverse vocabulary.

Lexical diversity may not have affected the quality of writing because lexical diversity describes how often different words are used but not how they are used or where they are used, or what the different words are. Participants who used a greater diversity of vocabulary might have made more mistakes in their usage, resulting in lower overall text quality, because of insufficient knowledge of the grammatical structures or logical relations relevant to the complex vocabulary being used. Conversely, participants who used a smaller range of vocabulary may have used these words with greater understanding and sophistication, resulting in higher text quality scores. Therefore, the lack of relationship between lexical diversity and proficiency levels might be explained by the higher levels of lexical diversity needed to minimise the amount of lexical repetition and maintain adequate language skills.

As discussed, the results revealed that lexical diversity may not be a defining characteristic of text quality, or at least not for this genre. English teachers in China tend to encourage students to focus on demonstrating a diversity of vocabulary in their writing. However, the results could indicate that it may be better for teachers to pay more attention to other dimensions besides lexical diversity in developing the written proficiency of their students. As explained, writing is a complex process that requires both linguistic and cognitive skills, as well as knowledge of the topic being written about. Teachers also need to focus on aspects such as coherence, organization, cohesion, style, grammar and audience awareness. Even in terms of vocabulary, lexical diversity is not the only dimension to be considered. Other dimensions include, but are not limited to, lexical accuracy, lexical

frequency and word association. Vocabulary knowledge is a complex construct that includes both breadth and depth (Anderson & Freebody, 1981). Vocabulary breadth refers to the size of a learner's vocabulary. Vocabulary depth reflects the quality of vocabulary use. As lexical diversity mainly relates to the breadth of vocabulary knowledge, it can only describe the range of words used in the compositions, but it cannot describe the quality of vocabulary use, for example, how different words are used or what the different words are. It is, therefore, possible that a text with a high degree of lexical diversity may not be rated highly, as it does not necessarily demonstrate depth of vocabulary knowledge.

The participants in Level 3 (the lowest level) on average produced the fewest types and tokens. It is possible that students at lower writing proficiency levels had difficulty in producing vocabulary. Vocabulary teaching could therefore be one starting point for teachers to help Chinese high school students who are less proficient in English to improve their writing quality.

This study investigated the relationship between lexical diversity and EFL writing proficiency levels for Chinese learners in an email writing task. The findings are specific to this L1 background and genre, but a limitation is that it is not possible to determine whether the lack of relationship between lexical diversity and EFL writing proficiency was related to language background or genre or an interaction between these two factors. Another limitation is the small sample size, meaning that there was lack of power to detect significant differences. The research focused on a sample of forty-five texts from three proficiency levels from one school. Further studies investigating larger samples and systematically investigating the effects of both language background and genre would enrich our understanding of the relationship between lexical diversity and writing. Moreover, the findings of this study were constrained in terms of the grade level of participants. All who took part in the study were in their last year of high school. It is unknown whether similar findings would have resulted from the inclusion of older or younger students. Finally, although the raters were experienced and reliable, there are still certain flaws in the NMET rating system. An issue is the

difficulty in distinguishing between the quality of two adjacent scores such as 15 and 16. Weaknesses such as this in rating systems may have negative effects on research outcomes (Weigle, 2004).

A strength of the study has been the use of two measures of lexical diversity, TTR and D. The utilisation of these two methods of measuring lexical diversity has enabled the current study to counteract potential problems associated with each method and thus to maximize the validity of the study.

CONCLUSIONS

This study has explored the relationship between lexical diversity and writing proficiency levels in Chinese high school students, using D and TTR to measure lexical diversity in an email writing task taken from a NMET practice test. The results indicated that the lexical diversity in email texts garded at higher proficiency levels was not significantly different from email texts garded at lower proficiency levels, and that the email texts displayed considerable variation in lexical diversity within the three levels of proficiency. Therefore, within this context, lexical diversity measures did not offer a basis for distinguishing the EFL writing proficiency levels of these Chinese students.

THE AUTHOR

Xuan Wang is a PhD candidate at the Faculty of Education and Social Work, University of Sydney. Her research interest is the relationship between vocabulary knowledge and writing quality in an EFL context. The findings presented in this article are based on her Dissertation which was completed in partial fulfilment of her Master of Education (TESOL) degree at the University of Sydney.

REFERENCES

- Alderson, J. C. (2005). *Diagnosing foreign language proficiency*. London: Continuum.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117), Newark, DE: International Reading Association.
- Arnaud, P. J. L. (1984). The lexical richness of L2 written productions

- and the validity of vocabulary tests. In T. Culhane, C. K. Bradley & D. K. Stevenson (Eds), *Practice and problems in language testing: Papers from the International Symposium on Language Testing* (pp. 14–28). Colchester, UK: University of Essex.
- Duran, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics*, 25, 220–242.
- Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4, 138–155.
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds), *Theoretical models and processes of reading* (pp. 928–950). Newark, DE: International Reading Association.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In R. B. Ruddell & N. J. Unrau (Eds), *Theoretical models and processes of reading* (pp. 1399–1430). Newark, DE: International Reading Association.
- Jarvis, S. (2002). Short texts, best fitting curves, and new measures of lexical diversity. *Language Testing*, 19, 57–84
- Klee, T., Stokes, S., Wong, A., Fletcher, P., & Gavin, W. (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language and Hearing Research*, 47, 1396–1410.
- Laufer, B. (2003). The influence of L2 on L1 collocational knowledge and on L1 lexical diversity in written expression. In V. Cook (Ed.), *Effects of the Second Language on the First* (pp. 19–31). Clevedon, UK: Multilingual Matters.
- Leki, I., & Carson, J. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, 28, 81–101.
- Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19, 85–104.

- Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development: Quantification and assessment*. New York: Palgrave Macmillan.
- McCarthy, P. M., & Jarvis, S. (2007). *vocd*: A theoretical and empirical evaluation. *Language Testing*, 24, 459–488.
- Meara, P., & Miralpeix, I. (2008). *D_Tools: The manual*. <http://www.lognostics.co.uk/>.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- Nation, P., & Heatley, A. (1996). *Range computer program*. English Language Institute, Victoria University of Wellington, Wellington.
- Nation, P., & Webb, S. (2011). *Researching and analysing vocabulary*. Boston, MA: Heinle, Cengage Learning.
- Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing*, 22, 545–565.
- Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing: An Interdisciplinary Journal*, 26, 45–65.
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.
- Read, J., & Nation, P. (2006). An investigation of the lexical dimension of the IELTS Speaking Test. *IELTS Research Reports*, 6, 207–231.
- Scardamalia, M., & Bereiter, M. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics, Vol. 2: Reading, writing, and language learning* (pp. 142–175). New York: Cambridge University Press.
- Schoonen, R., van Gelderen, A., Stoel, R. D., Hulstijn, J., & de Glopper, Kees. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning*, 61, 31–79.
- Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9, 27–

55.

Wimmer, G., & Altmann, G. (1999). Review article: On vocabulary richness. *Journal of Quantitative Linguistics*, 6, 1–9.

Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 31, 236–259.

APPENDIX 1: TASK INSTRUCTION**Guided Writing (Translated English Version)**

Suppose you are Li Hua from Xinhua Middle School. You plan to apply to an American university after you graduate from senior high school. According to the requirements of the university, you need to send a personal statement to the university. Please write an email letter and include the following information in your personal statement:

1. Self-introduction
2. Hobbies and interests
3. Reasons for application

Notes:

1. Write around 100 words
2. Provide details to make your writing coherent and cohesive
3. Provide an introductory and concluding sentence

APPENDIX 2: EVALUATION CRITERIA

First level (21–25 points)

Fully satisfies all the requirements of the task

- Fully presents the information
- Uses a wide range of vocabulary and grammatical structures
- Makes a few errors because of attempts to use more complex grammatical structures and more advanced vocabulary
- Uses cohesion effectively to make writing coherent

Fully achieves the anticipated goal of writing

Second level (16–20 points)

Satisfies the requirements of the task

- Presents the main information but omits one or two less important details
- Uses a variety of grammatical structures and vocabulary and satisfies the task requirements
- Uses grammatical structure and vocabulary basically accurately, with a few errors because of attempts to use more complex grammatical structures and more advanced vocabulary
- Uses simple cohesion to make writing coherent

Achieves the anticipated goal of writing

Third level (11–15 points)

Basically satisfies the requirements of the task

- Presents the main information but omits some information
- Uses a variety of grammatical structures and vocabulary and satisfies the task requirements
- Has a few errors that do not influence comprehension
- Uses simple cohesion to make writing coherent

Basically achieves the anticipated goal of writing

Fourth level (6–10 points)

Completes the requirements of the task inappropriately

- Omits some important information and writes something irrelevant content
- Uses limited grammatical structures and vocabulary
- Has a few errors related to grammatical structure and vocabulary that influence comprehension
- Uses less cohesion or lacks cohesion and coherence

Provides information but does not do so clearly

Fifth level (1–5 points)

Does not finish the requirements of the task

- Obviously omits main information, writes something irrelevant, and does not comprehend requirements
- Uses monotonous grammatical structures and limited vocabulary
- Has more errors related to grammatical structure and vocabulary that influence comprehension
- Lacks cohesion and coherence

Does not convey information to readers