Lexical Indicators of L2 Writing Performance

Kutay Uzun[1]

[1] Trakya University, Turkey / Contact: kutayuzun@trakya.edu.tr

Abstract

Performance in L2 writing is a complex phenomenon which encompasses complexity, accuracy and fluency. However, relevant literature indicates that lexical performance also plays a role in L2 writing performance as a whole. In this respect, the present study aims to find out which lexical indicators are related to L2 writing performance. Due to the correlational nature of the study, a quantitative research design was preferred. Analyses were performed on a corpus of 160 literary analysis essays written during a compulsorily taken English Literature course by 40 second year students of English Language Teaching at a public university in Turkey. Lexical Complexity Analyzer, which is a reliable software that produces numerical values for lexical performance indicators, was used for the analyses. L2 writing performance scores were assigned to each essay using a 6-point holistic rubric and the essays were re-scored 6 weeks after the first scoring for intrarater reliability. Following the reliability analysis, all variables were tested for the normality of distribution and essay scores were found to be non-normally distributed. Since this variable were to be present in all correlation analyses, Spearman's Rank Order Correlation Coefficient was calculated to see if there was a relationship between essay scores and each of the variables. After the tests of normality, relationships were sought between essay scores and lexical density, lexical sophistication, verb sophistication, number of different words, type/token ratio, lexical variation, verb variation, noun variation, adjective variation, adverb variation and modifier variation. The results showed that L2 writing performance was significantly correlated with lexical sophistication, adjective variation, adverb variation and modifier variation, however, the effect sizes of the significant correlations were too small, so the identified relationships were negligible. On the other hand, significant correlations with small effects were found between L2 writing performance and the number of different words, type / token ratio and verb variation.

Introduction

Writing is a crucial and functional skill for humans which requires a harmonious use of working memory, long-term memory and motor muscle skills (Burdick et al., 2013). The importance of the skill of writing, especially in academic contexts, arises from the fact that most of the assessment in those contexts take place in the written form since higher education contexts can be crowded in terms of class sizes (Alagözlü, 2012) and spoken forms of assessment may become virtually impossible because of the number of students. Also considering that one's mastery of writing skills is among the predictors of general academic success (Graham & Perrin, 2007), the significance of writing skills becomes even more visible, making the assessment of this skill equally crucial.

Even though it is possible to come across with varying approaches to the assessment of writing, it is seen that textual features are predominant in the process. In both qualitative and quantitative terms, the level of productivity and text quality is typically assessed in writing (Kim, Otaiba, Folsom, Greulich, & Puranik, 2014). In an earlier study, Wilkinson (1989) comprehensively lists the constructs to be assessed in writing as layout, organization, diction, content, logic, unity, vocabulary, accuracy, coherence, style and cohesion. In a similar manner, Wilson, Olinghouse, McCoach, Santangelo and Andrada (2016) suggest that grammar, syntactical accuracy, logical sequence and cohesion are among the constructs to be assessed in writing development. Using more comprehensive terms for similar constructs, Banerjee, Franceschina and Smith (2007) articulate syntactic complexity, grammatical accuracy, cohesion and lexical richness for the assessment of writing skills.

In the literature, it can be seen that most studies with respect to L2 writing skills place their foci on the syntactic components of L2 writing performance as the syntactic norms regarding the assessment of writing performance appear to be clearer (Polio, 2001). However, lexical richness, which encompasses the language use of a learner in terms of lexical density, lexical variation, lexical sophistication and vocabulary errors

should also be considered among the constructs which constitute L2 writing performance in valid and reliable terms (Lu, 2012). Among the components of lexical richness, lexical density is a well-known construct and it can be briefly defined as the ratio of content words against the total number of words (Ure, 1971). Lexical sophistication is related to the frequency of the words in a given text, defined as the number of less frequent or advanced-level words relative to the total number of words in a text (Read, 2000). Regarding word frequencies, Laufer and Nation's (1995) list of the first and the second most frequent 1000 words in English, as well as Xue and Nation's (1984) university word list are known to be commonly used for lexical sophistication analyses. A third construct within the domain of lexical richness is lexical variation, referring to the range of vocabulary as manifested in language use (Lu, 2012), typically measured as the number of *different* words in a text (Klee, 1992; Miller, 1991). Lexical variation can also be traced by means of the type-token ratio in a text, which is produced by comparing the word types in a text against word tokens, resulting in the word variation value (Skalicky, Crossley, McNamara & Muldner, 2017). A related construct is lexical word variation, which is measured as the proportion of lexical word types against the total number of lexical words (Casanave, 1994). With a similar calculation method, that is, by comparing the number of types against tokens limiting the computation to a target word form, verb variation (Harley & King, 1989), noun variation, adjective variation, adverb variation and modifier variation (McClure, 1991) can also be produced numerically. When those text-related definitions are taken into account, it seems necessary to study the lexical indicators of L2 writing performance along with the syntactic or grammatical ones to obtain a more thorough picture of the construct.

Within the Turkish higher education context, lexical indicators do not appear to have been extensively studied in relation to L2 writing performance, however, it is possible to come across with related studies. For instance, Köksal (2013) reports that undergraduate students find it difficult to improve themselves in terms of lexical density. In a similar manner, Şanal (2007) identifies a low level of lexical sophistication in a similar group of students. In terms of lexical variation, Bozdağ (2014) also identifies a low level among undergraduate students in Turkey and concludes that their texts typically involve recurrent choices, reducing variability. Şanal's (2007) findings

are parallel to those of Bozdağ (2014), indicating a low type-token ratio among undergraduate students. On the other hand, there appears to be no study with respect to lexical word variation, verb variation, noun variation, adjective variation, adverb variation and modifier variation within the same context.

As seen in the relevant literature, writing is a crucial language skill which also has a say in a learner's academic success. However, the assessment of this important skill is a complex process, which includes lexical as well as grammatical constructs to be evaluated. Even so, studies appear to lack sufficient contextual conclusions in terms of the lexically-based constructs and their relationship with L2 writing performance. In this respect, the study aims to fill a contextual gap by finding out which of the lexical indicators are related to L2 writing performance within the Turkish higher education context.

Purpose and Research Questions

Taking the gap in the literature into account, the present study aims to find out which of the lexical indicators described in the relevant literature may be correlated with L2 writing performance as manifested in essay scores.

The following research question was formulated for the aims of the study:

- Which lexical indicators are significantly correlated with L2 writing performance as manifested in literary analysis essay scores?

Methodology

Since the study was of a correlational nature, a quantitative design was preferred. According to Creswell (2012, 2014), quantitative designs can be used in true experiments and quasi-experiments as well as non-experimental designs such as causal-comparative or correlational designs, in the last one of which the researcher attempts to establish a potential relationship between two variables in numerical terms. When the aim of the study, which was to look for potential relationships between each one of the lexical indicators and L2 writing performance, was considered, a quantitative design appeared to have been the most suitable type of design.

A total of 160 literary analysis essays written during a compulsorily-taken English Literature course in the 2nd year of an English Language Teaching department of a public Turkish university were used as the corpus of the study. The literary analysis essays were 5-paragraph essays that were argumentative in nature, discussing how a particular theme or character in a literary work was dealt with by its author. The shortest essay within the corpus was found to have 57 words and the longest one had 845 words. The entire corpus was constituted by 48866 words with an average of 305 words per essay.

The essays were scored by the researcher, who was also the course instructor, using the Bauer & Kohut Argumentative Writing Rubric (Bauer, 2016). The rubric aims to evaluate argumentative writing performance in a holistic manner with a minimum score of 1 and maximum score of 6, taking Bloom's Taxonomy of Educational Objectives as its basis. In the original study, the rubric was found to be a valid and reliable one, which resulted in total agreement among 4 raters ($K = 1.00$, $p < .001$). In the present study, the rubric was initially studied in detail by the researcher and then the essays were skimmed for familiarization purposes. After skimming, the essays were read in detail for scoring. 6 weeks after the first scoring, 30% of the essays were scored once again by the researcher, producing a Cohen's Kappa intrarater reliability coefficient of .861 which was statistically significant ($p < .001$), indicating that the scoring of the essays was reliable.

In line with the relevant literature, the lexical variables dealt within the present study were lexical density, lexical sophistication, verb sophistication, number of different words, type-token ratio, lexical variation, verb variation, noun variation, adjective variation, adverb variation and modifier variation. For the computation of these variables, Lexical Complexity Analyzer which was developed by Ai and Lu (2010) and Lu (2012) was used. According to its developers, the Lexical Complexity Analyzer makes use of the Stanford Tagger (Toutanova, Klein, Manning & Singer, 2003) and MORPHA (Minnen, Carol & Pearce, 2001) which are both highly accurate and reliable pieces of software for lemmatization and part-of-speech tagging, making the software both valid and reliable (Ai & Lu, 2010; Lu, 2012).

In order to see the descriptive results with respect to the variables of interest, mean, standard deviation, median, minimum and maximum values for each variable were computed. To find out which correlation coefficient was to be used to meet the aims of the study, a Shapiro-Wilk's test of normality was initially run. The results indicated that the only variables within the context of the study that were normally distributed were lexical density ($SW = .986$, $df = 160$, $p = .116$) and lexical variation ($SW = .992$, $df = 160$, $p = .476$). All other variables, namely essay scores, lexical sophistication, number of different words, type-token ratio, verb variation, noun variation, adjective variation, adverb variation and modifier variation produced statistically significant Shapiro-Wilk coefficients ($p < .05$), indicating deviations from normality. Since all variables were to be tested for possible correlations with essay scores and essay scores were non-normally distributed, Spearman's Rank-Order Correlation Coefficient, which is a non-parametric correlation coefficient, was calculated for each variable. Effect sizes for the correlation coefficients were computed as r_s^2.

Findings

To find out the average level of the corpus with respect to each variable, descriptive analyses were initially run. The results are presented below in Table 1.

Table 1. Descriptive Findings of the Variables (N = 160)

Construct	*M*	*SD*	*Md*	*Min*	*Max*
Essay Score	3.54	1.47	4.00	1.00	6.00
Lexical Density	.53	.03	.53	.42	.63
Lexical Sophistication	.39	.08	.38	.10	.60
Verb Sophistication	.12	.07	.12	.00	.35
Number of Different Words	138.34	36.52	134.50	42.00	301.00
Type-Token Ratio	.46	.07	.45	.29	.70
Lexical Variation	.73	.12	.73	.41	1.00
Verb Variation	.19	.08	.18	.07	.57
Noun Variation	.54	.10	.54	.30	.90
Adjective Variation	.09	.03	.09	.02	.20
Adverb Variation	.06	.02	.06	.00	.12
Modifier Variation	.15	.04	.15	.08	.26

As seen in Table 1, the mean essay score within the corpus was 3.54 out of 6 ($SD = 1.47$) with a minimum of 1.00 and a maximum of 6.00. In addition, the mean was found to be 138.34 ($SD = 36.52$) in terms of the number of different words. According to the results, the values that were higher than 50% were lexical word variation ($M = .73$, $SD = .12$), lexical density ($M = .53$, $SD = .03$) and noun variation ($M = .53$, $SD = .03$). On the other hand, verb sophistication ($M = .12$, $SD = .07$), adjective variation ($M = .09$, $SD = .03$) and adverb variation ($M = .06$, $SD = .02$) were found to be particularly low in comparison to other values.

As mentioned in the methodology section, a Spearman's Rank-Order Correlation coefficient was calculated in relation to the essay scores for each one of the variables. Since there were 11 different variables, the correlational findings were reported in groups as those which did not have a statistically significant relationship, those which had negligible relationships due to the very low effect sizes despite the statistically significant probability value and those which had statistically significant relationships with the essay scores. The results in respect of the variables which produced no statistically significant relationship with the essay scores were given below in Table 2.

Table 2. Correlational Results with No Statistical Significance (N = 160)

Construct	Value	Lexical Density	Lexical Sophistication	Verb Sophistication	Lexical Variation
Essay Score	r	.013	-.116	.105	-.057
	p	.872	.145	.186	.473
	$r_s{}^2$	.00	.01	.01	.00

As seen in the findings, 4 of the lexically-based variables produced no statistically significant correlation with the essay scores. Among these, the Spearman's Rank-Order Correlation Coefficients were .013 for lexical density ($p = .872$, $r_s{}^2 = .00$), -.057 for lexical variation ($p = .473$, $r_s{}^2 = .00$), .105 for verb sophistication ($p = .186$, $r_s{}^2 = .01$) and -.116 for lexical sophistication ($p = .145$, $r_s{}^2 = .01$).

As mentioned previously, some of the variables produced statistically significant correlation coefficients with very small effect sizes, making the statistical significance negligible. These findings were reported below in Table 3.

Table 3. Significant Correlations with Very Small Effects (N = 160)

Construct	Value	Noun Variation	Adjective Variation	Adverb Variation
	r	-.194	-.184	-.158
Essay Score	p	.014	.020	.046
	r_s^2	.04	.03	.02

As shown in Table 3, 3 of the lexically-based variables had statistically significant correlations with the essay scores, however, the sizes of the effects of those correlations were too small to have practical implications. Those variables were noun variation (r = -.194, p = .014, r_s^2 = .04), adjective variation (r = -.184, p = .020, r_s^2 = .03) and adverb variation (r = -.158, p = .046, r_s^2 = .02). These variables were all found to have negative correlations with the essay scores and their correlation coefficients explained only 4% of the variance or less.

In addition to variables which had negligible or no relationship with the essay scores, some variables produced statistically significant correlation coefficients with small effects. The findings related to those variables were presented below in Table 4.

Table 4. Significant Correlations with Small Effects (N = 160)

Construct	Value	Number of Different Words	Verb Variation	Type-Token Ratio	Modifier Variation
	r	.455	.321	-.342	-.222
Essay Score	p	.000	.000	.000	.005
	r_s^2	.21	.10	.12	.05

As seen in the findings, 4 lexically-based variables were significantly correlated with the essay scores. According to the results, the number of different words (r = .455, p < .001, r_s^2 = .21) and verb variation (r = .321, p < .001, r_s^2 = .10) had weak and positive correlations with the essay scores. On the other hand, type-token ratio (r = -.342, p < .001, r_s^2 = .12) and modifier variation (r = -.222, p = .005, r_s^2 = .05), had a weak and negative correlations with the scores. It was also seen that all of these 4 variables produced small effects, explaining 21%, 12%, 10% and 5% of the variance respectively.

Discussion and Conclusions

The present study aimed to reveal which lexically-based constructs had statistically significant correlations with L2 writing performance in a corpus of short literary analysis essays written by 2nd year English Language Teaching students in a public university in Turkey. The findings indicated that while the corpus demonstrated higher levels of lexical word variation, lexical density and noun variation, it was also seen that the levels of verb sophistication, adjective variation and adverb variation were particularly low within it. According to the correlational results, lexical density, lexical sophistication, verb sophistication and lexical variation were not correlated with L2 writing performance. On the other hand, noun variation, adjective variation and adverb variation produced statistically significant correlations with negligible effect sizes. Lastly, the number of different words in a text, type-token ratio, verb variation and modifier variation values were found to have statistically significant correlations with L2 writing performance, indicating small effects.

The descriptive findings of the study, which showed that the corpus produced passable values only in terms of lexical word variation, lexical density and noun variation while all the other values were found to be in the lower half of the continuum, can be considered parallel to the findings of Şanal (2007) and Bozdağ (2014) as they also concluded that the levels of lexical sophistication and variation were low among Turkish undergraduate students. On the other hand, the findings appeared to be contradicting Köksal's (2013) and Bozdağ's (2014) some other findings in that while those studies reported low levels of lexical density and type-token ratio, it was seen that these values in the present study were among the higher ones. However, contextual differences of the studies may have accounted for the differences in the findings as Köksal's (2013) study was conducted in a French as a Foreign Language context. There, the main difference lies in the fact that English is started to be taught from very early ages unlike French in Turkey. Therefore, the participants of Köksal's (2013) study were highly likely to have been less proficient in French in comparison to how proficient the writers of the essays in the present study were in English, which may have resulted in the difference between the findings of two studies. On the other hand, even though Bozdağ's (2014) research context was similar, the results obtained were

different from those of the present study. In this respect, it can be said that the size of Bozdağ's (2014) corpus, which was twice as large as that of the present study, and the generic argumentative nature of the essays within may have accounted for the difference in the findings of two studies since the corpus of the present study was limited to literary analysis essays written in an argumentative style.

Correlation analyses revealed that noun, adjective and adverb variation were very weakly related to L2 writing performance producing very small effects, which were negligible (Cohen, 1988; Coe, 2002). In addition, lexical density, lexical sophistication, verb sophistication and lexical variation were found to have no relationship with L2 writing performance. In other words, in the texts written within the learning context of the present study, the variability of nouns, adjectives and adverbs as well as the ratio of content words and less frequent words to the number of words in the text did not appear to be related to L2 writing performance. However, analysis also showed that the number of words and verb variation were positively related to L2 writing performance. This meant that the more variability in the generic use of words and verbs demonstrated by the learners, the better performance was measured in their L2 writing. Even though some of the analysis results were not statistically significant, the variables which were so corroborated Lu's (2012) and Banerjee et al.'s (2007) suggestions that lexical richness was a part of L2 writing performance since a higher level of lexical variability, including verb variability, meant a higher essay score in the findings of the present study.

Analyses also resulted in conflicting findings in that type-token ratio and modifier variation were found to be negatively correlated with L2 writing performance. Even though writing performance is related to the number of words and verb variation, it was seen that the correlational findings indicated relationships towards opposite directions for type-token ratio and modifier variation. The reason why there was such a conflict in terms of the findings may have been that syntactically-based constructs, which were not constructs of interest within the present study, may have interfered with the essay scores within the corpus. Since the number of different words is affected by text length, writing fluency, which is also a syntax-oriented construct related to the length of a text (Polio, 2001), may have interfered with the scores, resulting in higher scores as text length increased. On the other hand, the finding which showed that lower

levels of type-token ratio and modifier variability resulted in a higher essay score may have signalled deviations from the actual essay topic or a lack of focus in the essays with higher levels of type-token ratio and modifier variability, resulting in lower essay scores. In other words, evaluation criteria which were related to the content and accuracy of the essays might have interfered with the scores and this interference may have caused a reduction in the scores due to issues with content even though the type-token ratio and modifier variability were relatively high in some of the essays.

Considering all the findings of the study, it can be concluded that lexical richness is, indeed, a part of writing performance and variability in the choice of words may result in a higher performance in L2 writing. On the other hand, it can also be inferred that content-related and syntactical measures such as writing fluency and grammatical accuracy may interfere with this relationship, therefore, utilizing solely lexical indicators for the measurement of writing performance may prove inaccurate as well as insufficient. For this reason and as already proposed in the relevant literature, lexical, syntactic, contextual and socio-discoursal features should be taken into account for the measurement of L2 writing performance.

As for instructional implications, the findings suggest that writing instruction should not be limited to syntactic components of writing performance and lexical components should also be taken into account to produce better outcomes among learners. Moreover, the findings may also be interpreted to be suggestive of a necessity to provide written corrective feedback that is directed at lexical richness in addition to feedback that focuses on grammatical accuracy, which is more typical for L2 writing contexts.

The limitations of the study should also be considered in the interpretation of the findings. The corpus used in the study was limited to short literary analysis essays written in an undergraduate English as a Foreign Language context, therefore, a different or more generic corpus may produce different results in similar analyses. In addition, the potential interference of different constructs such as syntactic complexity or grammatical accuracy with the findings cannot be conclusively argued for as they were not investigated within the context of the study. For this reason, further research can be directed towards multivariate predictive analyses using all the constructs

possible or controlling for some of the constructs to identify potential relationships or interferences.

Notes on the contributors

Kutay Uzun (PhD) is a lecturer in the Department of English Language Teaching at Trakya University, Edirne, Turkey. His research interests are writing in EFL and educational data mining.

References

Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Diaz-Negrillo, N. Ballier, & Paul Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249-264). Amsterdam/Philadelphia: John Benjamins.

Alagözlü, N. (2012). English as a foreign language cul-de-sac in Turkey. *Procedia – Social and Behavioral Sciences, 47*(2012), 1757-1761.

Banerjee, J., Franceschina, F., & Smith, A. M. (2007). *Documenting features of written language production typical at different IELTS band score levels (IELTS Research Reports Volume 7)*. Australia & UK: IELTS Australia and British Council.

Bozdağ, F. Ü. (2014). *Lexical verbs in academic writings of Turkish learners of English as a second language: A corpus-based study*. Unpublished Master's thesis, Çukurova University, Adana, Turkey.

Burdick, H., Swartz, C., Stenner, J., Fitzgerald, J., Burdick, D., & Hanlon, S. (2013). Measuring students' writing ability on a computer-analytic developmental scale: An exploratory validity study. *Literacy Research & Instruction, 52,* 255-280. doi: 10.1080/19388071.2013.812162

Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing, 3*, 179–201.

Coe, R. (2002). It's the effect size, stupid! What effect size is and why it is important. Paper presented at the Annual Conference of the British Educational Research

Association, University of Exeter, England. Available at https://www.leeds.ac.uk/educol/documents/00002182.htm

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Hillsdale: Lawrence Erlbaum.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.)*. Upper Saddle River, NJ: Merrill.

Creswell, J. W. (2014). *Research Design Qualitative, Quantitative and Mixed Methods Approaches (4th ed.)*. Thousand Oaks, CA Sage.

Graham, S., & Perrin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445-476.

Harley, B., & King, M. L. (1989). Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition, 11*, 415–440.

Kim, Y., Otaiba, S. A., Folsom, J. S., Greulich, L., and Puranik, S. (2014). Evaluating the dimensionality of first-grade written composition. *J. Speech Lang. Hear. Res. 57*, 199–211. doi: 10.1044/1092-4388(2013/12-0152)

Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders,12*, 28–41.

Köksal, O. (2013). The impact of the keyword method on vocabulary learning and retention in preparatory French classes in higher education. *International Journal of Academic Research, 5*(5), 393-400. doi: 10.7813/2075-4124.2013/5-5/B.61

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*, 307–322.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal, 96*(2), 190-208. doi: 10.1111/j.1540-4781.2011.01232.x

McClure, E. (1991). A comparison of lexical strategies in L1 and L2 written English narratives. *Pragmatics and Language Learning, 2*, 141–154.

Miller, J. F. (1991). Quantifying productive language disorders. In J. F. Miller (Ed.), *Research in child language disorders: A decade of progress* (pp. 211–220). Austin, TX: Pro-Ed.

Minnen, G., Carroll, J., & Pearce, D. (2001). Applied morphological processing of English. *Natural Language Engineering, 7,* 207-223.

Polio, C. (2001). Research methodology in second language writing: The case of text-based studies. T. Silva and P. Matsuda. (Eds.) *On second language writing.* (p. 91-116). Mahwah, NJ: Erlbaum.

Read, J. (2000). *Assessing vocabulary.* Oxford: Oxford University Press.

Skalicky, S., Crossley, S. A., McNamara, D. S., & Muldner, K. (2017). Automatically identifying humorous and persuasive language produced during a creative problem-solving task. In *FLAIRS 2017 - Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference* (pp. 282-287). AAAI Press.

Şanal, F. (2007). *A learner corpus-based study on second language lexicology of Turkish students of English.* Unpublished PhD Dissertation, Çukurova University, Adana, Turkey.

Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 252-259). Edmonton, Alberta, Canada: The Association for Computational Linguistics.

Ure, J. (1971). Lexical density: A computational technique and some findings. In M. Coultard (Ed.), *Talking about text* (pp. 27–48). Birmingham, England: English Language Research, University of Birmingham.

Wilkinson, A. (1989). Assessing language development: The creditor project. In A. Freedman, I. Pringle, & Yalden (eds.), *Learning to write, First language/second language* (pp. 67-70). New York: Longman, Inc.

Wilson, J., Olinghouse, N. G., McCoach, B., Santangelo, T., & Andrada, G. N. (2016). Comparing the accuracy of different scoring methods for identifying sixth graders at risk of failing a state writing assessment. *Assessing Writing, 27*, 11–23. doi: 10.1016/j.asw.2015.06.003

Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication, 3*, 215–229.