Effect Sizes of Formulaic Sequences Used in English Writing

Xuefeng Qin, Yan Wu, Dongying Yang
College of Foreign Languages
North China University of Science and Technology
Tangshan, China
qxf0627@163.com

Abstract—The results of Null Hypothesis Significance Testing can only answer the question whether there is an effect. Effect sizes are commonly used to evaluate the magnitude of the effect and practical significance. Taking eight primary studies as an example concerning the effect of formulaic sequences used in English writing, this paper calculated and interpreted effect sizes and the practical significance of each study. Effect sizes have potential advantages in evaluating practical significance in foreign language research.

Keywords—effect sizes; practical significance; formulaic sequences; English writing

I. INTRODUCTION

During recent years, the effect of formulaic sequences has been recognized and researched on English writing of English learners of Chinese (Guo & Mao, 2010; Chen, 2015). An increasing number of empirical studies suggest that the utilization of formulaic sequences can improve the quality of English writing in different aspects such as accuracy, fluency and appropriateness of language use (Gai, 2010; Wu, 2014). These findings, however, mainly derived from the results of the Null Hypothesis Significance Testing (NHST), which has long been criticized due to the fact that statistical results are heavily influenced by the number of samples; thus it follows that statistical significance doesn't mean practical significance in the real world (Ellis, 2010; Zheng *et al*, 2011). To supplement a null hypothesis significance test, effect sizes are strongly advocated to evaluate the magnitude of research effect (Borenstein *et al*, 2009; APA, 2010, p. 30; Bao & Xi, 2010).

By reviewing primary empirical studies in the academic journals in China's mainland concerning the effect of formulaic sequences used in English writing , no effect sizes are found calculated and reported. The main reason is probably that scholars and researchers are not familiar with the meaning and calculation of effect sizes. So the aim of this study is to introduce the concept, meaning and calculation of effect sizes by using the data reported in some of primary studies with the hope of raising awareness of effect sizes in foreign language research in China's mainland.

II. STATISTICAL SIGNIFICANCE, PRACTICAL SIGNIFICANCE, AND EFFECT SIZES

Null Hypothesis Significance Testing can only answer the question whether there is an effect, not answer how much the effect it can produce (Cumming, 2012;Lu, 2003). In general, researchers believe that if the probability of statistical results is less than the given significance level (for example $\alpha = .05$), the null hypothesis is rejected and the alternative hypothesis should be accepted, thus drawing a conclusion that research results have significant effect. However, these findings only indicate the existence of research effects, but do not answer the question how much the effect it is, and whether there is practical significance in pedagogy, psychology and other specific research areas. In other words, the statistical significance of NHST results doesn't mean practical significance of research results. In contrast, if the NHST results don't match with statistical significance, they don't mean no practical significance (Ellis, 2010; Cumming, 2014). There is always confusion between statistical significance and practical significance.

Because the results are deeply affected by sample sizes, there are two phenomena in the NHST results. On one hand, in case research samples are relatively large, even if differences or relations are tiny, these findings also can present statistical significance with little practical significance. On the other hand, if research samples are too small in spite of larger differences or relations, these findings can't present statistical significance, but practical significance in reality can't be neglected.

To objectively evaluate the magnitude of practical significance in research effects and eliminate the influences of sample sizes on research effects, statisticians (Kirk, 1996; Lipsey & Wilson, 2001) propose the concept of effect size, which is used for calculating practical effect or magnitude of research effects. In the unit of standard deviation, effect sizes (standard score Z-score) are independent of sample sizes, which can be used to compare the sizes among the results of different measurement units. According to the diversity of research design in such areas as sociology, pedagogy and psychology, there are two kinds of common effect sizes. One is d family effect size for evaluating differences, which mainly refers to standardized mean deviation; the other is r family effect size evaluating strength of association, which

is defined as correlation coefficient. According to the standard for effect sizes, Cohen (1988) came up with guideline for measuring effect sizes on the basis of behavior science research. For d family effect size in between-group design, 0.2, 0.5, 0.8 represents small, middle and large effect size respectively; for r cluster effect size,0.10, 0.30, 0.50 represents small, middle and large effect size respectively. Plonsky and Oswald（2014）put forward the reference standard for evaluating effect size in second language research. For d family effect size in between-group designs, 0.40, 0.70, 1.00 represents small, middle and large effect size respectively; for d family effect size in within-group designs, 0.60, 1.00, 1.40 represents small, middle and large effect size respectively; for r family effect size,0.25, 0.40, 0.60 represents small, middle and and large effect size respectively.

Although there is some difference in size scale between the above two guidelines due to the different research fields, d family effect sizes are standard scores based on standard deviation. Effect sizes in different researches can be explained in accordance with normal distribution table. Normal distribution table can reflect areas or cumulative percentage when normal curve is lower than certain standard score. For example, if one effect size of a between-group design is 0.2 (the mean of the experimental group is 0.2 standard deviation higher than that of the controlled group), normal distribution table can show that 58% of the samples in the controlled group are lower than the mean of experimental group. The correlation coefficient of two variables, r, can be used as effect size in itself. The squared value of r (R^2 is called determination coefficient) can be explained as the shared amount of variability between the two variables. For example, if the correlation coefficient r between two variables in one study, also called effect size, is 0.3, its squared value is 0.09 accordingly, which means shared amount of variability between the two variables is 9%, or that the variation of one variable can explain 9% of variation of another variable. It is noted that, Cohen and other scholars advised researchers not to regard their guideline as doctrine and to explain effect size in accordance with specific research areas and problems, otherwise it will draw a wrong conclusion.

III. THE STUDY

A. Research aim

This research is to calculate and explain the effect size of some published primary studies based on the data they presented which didn't present effect sizes.

B. Data Collection

This research selected from Chinese core journals of the foreign language research eight published papers about empirical research of effect of lexical chunks used in English writing. The research used between-group experimental design, and statistical method was independent samples t-test. To calculate every effect sizes, the research extracted data including the sample size n_e, n_c of experimental groups and controlled groups, mean m_e, m_c, standard deviation s_e, s_c and significant value p (See table 1)

C. Research Instrument

The research used the online ES calculator[a] designed by Wilson to calculate effect size d of every research, namely Standardized Mean Difference[b]. The user can calculate effect sizes just by inputting the sample sizes of experimental groups and controlled groups, means and standard deviations on-line.

IV. RESULT AND DISCUSSION

The purpose of this research is to calculate the effect size for each of the eight published literature. Table 1 presents the data of the eight empirical research of effect of lexical chunks used in English writing. The results of all eight NHST reached statistical significance level, thus it prove that lexical chunks are effective in improving English writing. However, the results of these researches only answer the question whether there is an effect, but not how much the effect it is. From the effect sizes presented in table 1, we can draw the conclusion that the use of lexical chunks plays the positive role on English writing. Because the practical significance of each research is not the same, there is a big difference among the researches. According to the guideline of Cohen(1988), the research effect size (d=0.44) of Guo Yueqin and Wu Xiujuan and the research effect size(d=0.54) of Xu Jun and Huang Yonghua close to the middle effect size, the left six research effect sizes are all large effect sizes. According to the guideline of Plonsky and Oswald (2014) in second language research field, the research effect size (d=0.44) of Guo Yueqin and Wu Xiujuan and the research effect size (d=0.54) of Xu Jun and Huang Yonghua belong to small effect size. The effect size of Chen Donglan (2015) is between middle size and large size, and the left five effect sizes are all large effect sizes. Although the two guideline are different, they both indicate statistical significance that lexical chunks can make in improving English writing, but practical significance in this research field.

For a more visual explanation and understanding the significance of effect size in practical teaching we can give interpretation from two aspects by consulting the normal distribution table. On one hand, the percentage of controlled group which is lower than the mean of experimental group. On the other hand, Rank of person in a controlled group who would be equivalent to the average person in experimental group (Coe, 2002) (see table 1). For example, the effect size of Chen Donglan (2015) in the research of the use of lexical in improving English writing is 0.81, which indicates that 79% students' scores (39 students) of controlled group are lower than the mean (10.78) of experimental group, and this average ranks 11th in 50 students of the controlled group. The effect size of Deng Liming and Wang Xiangyun (2007) in the research of the use of lexical in improving English writing is

---

a. The website of Standardized Mean Difference of Wilson: http://cebcp.org/practical-meta-analysis-effect-size-calculator/.

b. The calculation formula of (d) is: $d = \frac{m_e - m_c}{s_p}$,in this formula, s_p refers to the polled standard deviation of the experimental group and the controlled

group, namely, $s_p = \sqrt{\frac{(n_e-1)s_e^2+(n_c-1)s_c^2}{n_e+n_c-2}}$. m_e, m_c, n_e, n_c, s_c and s_e refer to the mean, sample size and standard deviation of the experimental group and the controlled group.

TABLE 1: STATISTICS SUMMARY ON THE RESEARCH FOR THE EFFECT OF LEXICAL CHUNKS USED IN ENGLISH WRITING

	Experimental group			Controlled group						
	n	m	s	n	m	s	p	ES	Percentage[a]	Rank[b]
1	50	10.78	1.39	50	9.71	1.26	.000	0.81	79%（39）	11
2	49	102.39	11.02	53	89.19	11.07	.000	1.19	88%（46）	7
3	30	14.16	4.63	30	12.25	4.14	.018	0.44	67%（20）	10
4	31	25.07	2.07	33	21.85	2.99	.000	1.25	89%（29）	4
5	36	9.97	2.46	37	8.59	2.61	.023	0.54	71%（26）	11
6	50	10.33	1.46	50	7.98	1.75	.000	1.46	93%（46）	4
7	39	13.05	1.47	32	11.41	1.39	.000	1.14	87%（27）	5
8	37	6.86	0.69	34	4.78	0.71	.025	2.97	100%（34）	1

Notes: 1. Chen (2015); 2. Wu (2014); 3. Guo and Wu (2013); 4. Lou (2011); 5.Xu and Huang (2011); 6. Xiao (2011); 7. Gai (2010); 8. Deng and Wang (2007). The

The author has rounded up part of the data before keeping double digits that follow the decimal point for easy calculation..

a: percentage of controlled group who would be below average person in experimental group

b: Rank of person in a controlled group who would be equivalent to the average person in experimental group

2.97, which indicates that 100% students' scores of controlled group are lower than the mean scores (6.86) of experimental group, which ranks the first and it is higher than all the 34 students' scores of controlled group. Through the two interpretations reader can have more practical explanation and understanding of the practical significant of the use of lexical chunk in improving English writing.

The reasons leading to the research effect sizes different are related with many factors. Firstly, the characteristics of research samples are different. For example, some samples are freshmen and others are sophomore, and some samples are English majors and the others are non-English majors. Secondly, the duration of research varies. For example, some research lasted one semester and the others lasted two semesters. Thirdly, research tools are different. The requirements of researches in such aspects as the content and the difficulty levels of writing are inconsistent in the later of the experimental testing, so writing time may also different. Fourthly, the contents of lexical chunk teaching are different. It shows mainly the sources of lexical chunks in the experiment group are different, such as the different textbooks and so on. Finally, lexical chunks teaching methods are different. For example, in some research experimental group and controlled group share one teacher, but in other research the teacher of the two groups is different. And in some research the discussion-based teaching is used and in other research the autonomous learning method is adopted. The impact made by these factors in research can be regarded as moderator variable in meta-analysis to explore the source leading to system variation (Ellis, 2010; Cumming, 2012), thus we can have more comprehensive, scientific understanding and evaluation the real impact and significance that lexical chunks made on English writing.

V. CONCLUSION

Based on probability, NHST can only answer the question whether there is an effect, while independent of sample sizes effect sizes are statistical indices to evaluate the true size of research effects, and because the effect sizes are standard scores in standard deviation, it can be used to compare the true effect of the same research questions among different research design research. Taking eight researches as example, this paper introduced and presented

the effect sizes and practical significance of lexical chunks used in English writing to raise awareness of effect sizes in foreign language research in China's mainland and improve research level and quality. It can provide scientific basis to integrate the research effect of all single researches of the same research questions by using meta-analysis and to explore influence of moderators on research effects in research design.

REFERENCES

[1] M.H. Borenstein., L.V. Hedges, J.P.T. Higgins, and H.R. Rothstein, Introduction to Meta-Analysis. Hoboken, NJ: Wiley, 2009.

[2] R. Coe, "It's the effect size, stupid: What effect size is and why it is important", 2002, accessed from www.leeds.ac.uk/educol/documents/00002182.htm.

[3] J. Cohen, Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum. 1988.

[4] G. Cumming, "The new statistics: Why and how". Psychological Science, 2014, vol. 25, pp. 7–29.

[5] G. Cumming, Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis, New York: Routledge. 2012.

[6] P. D. Ellis, The Essential Guide to Effect Sizes: Statistical Power, Meta-analysis, and the Interpretation of Research Result, Cambridge, England: Cambridge University Press, 2010.

[7] R. E. Kirk, "Practical significance: a concept whose time has come", Educational and Psychological Measurement, 1996, vol. 5, PP. 746-759.

[8] M.W. Lipsey, and D. B. Wilson, Practical meta-analysis, Thousand Oaks: SAGE, 2001.

[9] L. Plonsky, and F. L. Oswald, "How big is "Big"? Interpreting effect sizes in L2 research", Language Learning, 2012, vol. 64, pp. 878-912.

[10] G. Bao, and Y. Xi, "Statistical Significance Testing Problems and Reflection", Joumal Nanjing Institute of Technology (Social Science Edition), 2010, vol. 10, pp. 27-32.

[11] D.L. Chen, "The Application of the Lexical Approach Driven by Language Awareness to the Teaching of College English Writing", Foreign Language Research, 2015, vol. 183, pp. 112-115.

[12] L.M. Deng, and X.Y. Wang, "Recitation Input for the Proficiency Development of L2 Chinese Student Writing", Foreign Language Education, 2007, vol. 28, pp. 52-56.

[13] S.H. Gai, "Empirical study of lexical chunk acquisition in the Zone of Proximal Development: An Empirical Study Based on Scaffolding Instruction", Foreign Language and Their Teaching, 2010, vol. 5, pp. 68-72.

[14] X.Y. Guo, and H.M. Mao, "An Empirical Research on Improving English Writing Ability by Lexical Chunks Teaching", Shandong Foreign Language Teaching Joumal, 2010, vol. 3, pp. 52-59.

[15] Y.Q. Guo, and X.J. Wu, "On the Application of CLIL in College English Writing", Foreign Language and Literature, 2013, vol. 29, pp. 166-169