

Received: October 27, 2017

Revision received: May 2, 2018

Accepted: May 13, 2018

Copyright © 2018 EDAM

[www.estp.com.tr](http://www.estp.com.tr)

DOI 10.12738/estp.2018.5.017 • October 2018 • 18(5) • 1160-1169

*Research Article*

## Research on the Effectiveness of College Student English Writing Teaching Based on Data-Driven Learning

Lidan Mao<sup>1</sup>  
Hebei Agricultural  
University

Yang Liu<sup>2</sup>  
Hebei Agricultural  
University

Mingjie Zhang<sup>3</sup>  
Hebei Agricultural  
University

### Abstract

Data-driven learning is a kind of new teaching mode under which language library is applied to language teaching. Featured by discovery learning, diversified learning process, media learning materials and instructional learning style, it can better stimulate students' learning initiatives, improve their self-learning ability and increase their learning efficiency. As the English writing level of college students is the most important indicator of students' subjective initiative in learning and their own English proficiency, this paper applies data-driven learning (DDL) in college students' English writing teaching in combination of the process writing approach. It selects 52 college students in a normal college in Hebei Province as the test subjects, and then through detailed experimental design and the pre-test, mid-test and post-test, discusses the effectiveness of data-driven learning in college English writing teaching. At last, through a questionnaire survey, this paper shows the students' attitudes towards the new data-driven learning method. This paper provides certain theoretical guidance to the improvement of the English writing level of college students.

### Keywords

Data-Driven Learning • Corpus • English Writing • Process Writing Approach

<sup>1</sup>**Correspondence to:** Lidan Mao (MA), College of Foreign Languages, Hebei Agricultural University, Baoding 071000, China. Email: mia\_mao@163.com

<sup>2</sup>College of Foreign Languages, Hebei Agricultural University, Baoding 071000, China. Email: mooncitylaura@hotmail.com

<sup>3</sup>College of Foreign Languages, Hebei Agricultural University, Baoding 071000, China. Email: zmj0312@outlook.com

Currently, computer technology, multimedia technology and network technology are developing very rapidly, which greatly influencing classroom are teaching methods and students' learning approaches. Diversified information, books, images, audio-visual data, network and so on have injected new elements into language teaching, and the absorption and refinement of these elements requires students to take more initiative to participate in learning. Modernized language teaching is not only limited to the knowledge transfer between teachers and students (Cazorla, 2001); more importantly, it is the positive interactions between students and English learning materials, and in the English teaching process, teachers play more of a role to stimulate students' abilities of self-analysis and problem solving.

Data-driven learning starts from corpus learning. It allows students to use indexing software like the corpus to inquire relevant language knowledge for learning and lets knowledge drive learners to improve their own language levels (Turney and Littman, 2005). Corpus presents a large number of examples sentences with real-world contexts in the form of data or concordance to learners, forming a learning environment that helps learners concentrate, enhance their memories and utilize contexts to acquire semantics and summarize the patterns. In the course of learning, it gradually makes learners aware of observing linguistic phenomena, analyzing and summarizing language features, and in this way, strengthens their language abilities. Thanks to all these functions of corpus, data-driven learning guides students to observe, discover and extract language usages and exploit grammatical and application characteristics from the real contexts (Aleahmad, Aleven & Kraut, 2009). The combination of data-driven learning and modernized language teaching can increase students' enthusiasm and participation in the English learning process, and such application has been gradually promoted in English language teaching.

With data-driven learning as the main learning approach, this paper applies the computer corpus technology in the English writing teaching for college students. In order to perform an experiment, it selects 52 college students in a normal college in Hebei as the test subjects, and then through detailed experimental design and the pre-test, mid-test and post-test, it compares and analyzes the English writing levels of the experimental class and the control class. At last, through a questionnaire survey, this paper summarizes the students' attitudes towards the new data-driven learning method. This paper demonstrates the effects of data-driven learning on English writing teaching with an example, which provides certain theoretical guidance to the promotion of data-driven learning.

## Theoretical Basis

### Constructivism theory

The constructivism theory emphasizes the knowledge construction process generated in people's consciousness in the process of learning. Unlike traditional learning, constructive learning does not transfer information from teachers to students, but rather relies on the students' active knowledge construction. When receiving new knowledge, the learner conducts self-cognition of knowledge construction, or changes the previous knowledge structure, or ignores the new knowledge (Lughofer and Kindermann, 2010). Constructive learning can activate students' metacognition, self-analysis, and introspection.

The constructive learning theory focuses on the authenticity and diversity of knowledge. Down to language learning, diverse realities and relevant context are the important foundation for constructivist language learning. Through the active and continuous construction and reconstruction of language learning, learners improve their own language skills.

### Data-driven learning

The data-driven learning theory follows the constructivism theory. By making full use of the computer network and corpus resources, it changes the traditional cramming teaching model, and by giving real context, it provides learners with the native language learning environment and helps students learn and solve problems by themselves (Solomatine, Maskey & Shrestha, 2010).

Corpus is the main tool and means for data-driven learning. It is a collection of language texts, but it is also different from texts, as shown in Table 1.

As shown in the table, different from the traditional texts which appear as whole paragraphs, corpus is based on single-line, separate contexts, which has formed a new way for learners to understand and learn the target language (Peng, Aston, Gunn, Liou & Ashburner, 2008).

Table 1  
The Comparison of Text and Corpus

Text	Corpus
Read the article as a whole	Read the article by paragraph
Horizontal reading	Portrait reading
Read only once	Read repeatedly
Personal will	Social practice sample
Language examples	Deeply understand the language
Coherent communication project	Non-coherent communication project

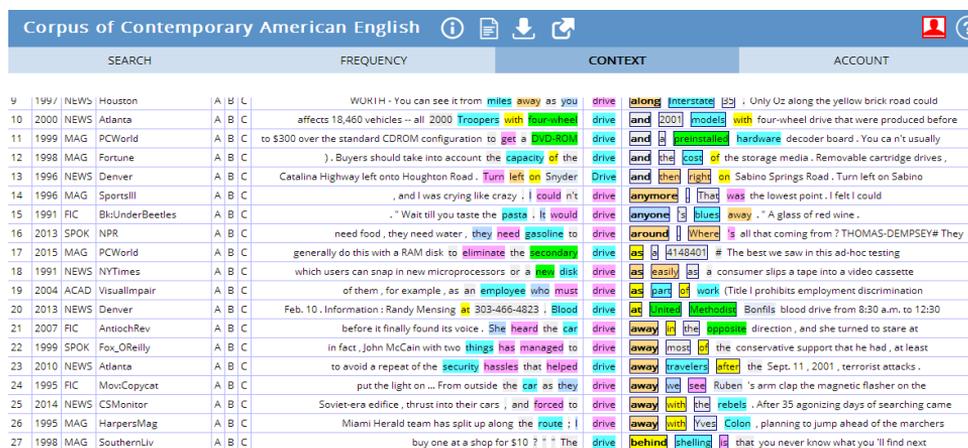


Figure 1. The retrieval result of “drive” in KWIC

For corpuses, the most widely used indexing method is the “key word in the context” (KWIC). Figure 1 shows the search of the key word “drive”. As shown in the figure, the example sentences containing “drive” in relevant contexts are presented in the search results (Talvensaaari, 2008; Talvensaaari, Laurikkala, Juhola & Keskustalo, 2007).

Corpus concordance not only provides rich contexts for choices of words, but also helps learners find out the small differences of the words in the contexts so that they will know how to collocate them. Less dependence on teachers and textbooks has brought about a growing learning initiative (Resnik & Smith, 2002).

Data-driven learning also has three advantages:

**Authentic texts.** The authenticity of corpus provides an authoritative data base for data-driven learning. The data collection of real texts provides students with the opportunity to carry out habitual communication in English, and the real language materials also help learners lay a solid foundation for gaining language proficiency.

**Inductive learning.** Inductive learning is a teaching model that is completely different from the cramming method. Students need to explore the language patterns and learn grammar, vocabulary, discourse and other knowledge under the guidance of teachers, allowing students to have a kind of exploratory language learning experience and stimulating their enthusiasm to enhance their language skills on their own.

**Independent learning.** The openness of the computer network allows independent learning to expand in both time and space. Students can use corpuses and other online learning opportunities to complete their self-learning according to their own levels and progress, without the interference from teachers or others.

### Process approach of English writing

English writing has always been a difficult part in Chinese students’ English learning. Many students have not grasped the basic process of writing and complete their English writing by simply piecing words together. As a result, it is very difficult for them to improve their writing quality (Espejo, Svalgaard & Dyer, 2006). The process approach of writing has its strict steps. Usually there are six steps: (1) brainstorming; (2) planning; (3) drafting; (4) editing; (5) revising; and (6) finalizing.

This paper takes the “Advantage of the world wide web” as an example to show the specific requirements for each step.

In the brainstorming stage, a student may come up with some words like “easy research”, “more examples” and “wide range of source”. He/she then records these ideas on his/her draft. At this time, applying a corpus can give him/her some inspirations. For example, if he/she searches “world wide web” in the corpus, he/she will see how “world wide web” is applied on different occasions, like in speeches, novels and magazine article, as shown in Figure 2. The student can select the types according to his/her needs, to gain some inspirations for writing (Modhish, 2005).



Figure 2. Retrieval result of “world wide web” in different sections

In the planning stage, the student can delete and organize these words to form a general framework and pivots for his/her article. Then the student begins to write according to the plan, looks for the pivots in the article and gradually completes the draft. Most of the first drafts need to be revised. The process of revising, the student needs to be clear about “what to write” and “how to write”, and revises and edits the article according to the proofreading comments from teachers and peer students (Inanoglu & Young, 2009). After completing drafting, revision and corrections and ensuring the grammar and fluency of the whole text, the student can finalize the article.

## Experimental design for effectiveness of college English writing teaching based on data-driven learning

### Research subjects

A total of 52 students from a normal college in Hebei are selected as the subjects, whose English skills are at a moderate level. The 52 students were divided into two classes – the experimental class and the control class, respectively. The data-driven method was adopted in the English writing teaching in the experimental class while the traditional English writing teaching method was adopted in the control classes (He, 2016).

### Research tools and data-driven writing teaching content

**Research tools.** This paper selects mainstream English corpora to provide corpus guidance to students in English writing. The main online free corpora include: British National Corpus (BNC), English-Chinese Parallel Concordancer and Corpus of contemporary American English (COCA).

**Teaching content.** Corpus training in data-driven learning mainly includes the following four parts:

(1) Searching lexical collocations. There are differences between English and Chinese in terms of semantics and expression. If we translate them bluntly, there might be inappropriate collocations. For example, “大马力汽车” should be “powerful car” rather than “strong car”; and “大风” should be “strong wind” rather than “big

wind”. With the vocabulary use frequency search function of the corpus, students can search unfamiliar lexical collocations to help them understand how to use and collocate words correctly (Narita, Kurokawa & Utsuro, 2003).

(2) Searching example sentences or common sayings. In writing, common sayings and idioms work as transitions, and more importantly, they are the refinement and distillation of the content. Searching common sayings and idioms in the corpus ensures students can use the idioms authentically, which is of great significance to improving their English writing skills.

(3) Comparing different ways of thinking and experiencing semantic prosodies. The ways of thinking are quite different between Chinese and English expressions. For example, in Chinese, an address is written in the sequence from the largest location to the smallest one, and in English, it is the reverse. Chinese people like to use verbs in their expressions while the Westerners like to use nouns. For example, it would be very Chinglish if “他能吃能睡” is translated into “he can eat and sleep”, because in English, the usual expression is “He is a good eater and sleeper”.

(4) Learning polysemous words in concordance. In English, polysemous words are used as frequently as those in Chinese. Native speakers have much better grasp of their own language than foreign language learners. Therefore, the concordance in the corpus can help foreign language learners learn and compare the usages of polysemous words in authentic contexts (Tong, 1989).

The above corpus training and the process writing approach are incorporated into the writing teaching in the experimental class.

### **Research objectives and steps**

**Research objectives.** 1. Whether the English writing teaching based on data-driven learning is more effective than the traditional writing teaching; 2. Whether the English writing teaching process based on data-driven learning goes smoothly; 3. What are students’ attitudes towards data-driven learning

#### **Research steps.**

(1) Pre-test. The pre-test score is the criterion for the selection of subjects. According to the pre-test results, 52 students of moderate English level are selected and assigned to the experimental class and the control class, respectively. The pre-test score is also the criterion for data evaluation and comparison after the end of the experiment.

(2) Stage-1 teaching. The stage lasts for a month, designed to allow the experimental comparison of data-driven teaching and traditional teaching. In the teaching process, the titles of essays, writing requirements and evaluation criteria are the same in the experimental class and the control class.

(3) Mid-test and immediate interview. After a month, the English writing scores of the experimental class and the control class are compared, and at the same time, the mid-test scores of the experimental class are also

compared with the pre-test scores. In the immediate interview, students are asked about the problems they have encountered in data-driven learning. The results are summarized and sorted.

(4) Stage-2 teaching. After the end of the mid-test, the second stage of teaching begins, where the shortcomings in the data-driven learning are improved according to the problems fed back by the students. The Stage-2 teaching continues until the end of the semester.

(5) Post-test. After the end of the semester, the English writing scores of the experimental class are compared with those of the control class, and with the mid-test scores of the experimental class. Teaching results and problems are analyzed based on the data.

(6) Questionnaire survey. The overall attitude of the students towards data-driven learning is obtained through the questionnaire survey.

## Experimental design for the effectiveness of data-driven English writing teaching in college

### Pre-test scores

The English writing pre-test results of the experimental class and the control class were analyzed, and at the same time the significance test of difference was carried out, as shown in Table 2:

Table 2

*The Pre-Test Score of Test Group and Control Group Comparison Char*

	Number	Mean	Std.Deviation	P
Test group	26	11.4231	3.3157	0.079
Control group	26	11.4808	3.1838	

It can be seen from the table that the English test scores of the experimental class and the control class were basically the same.  $T=0.079>0.05$ , indicating that before the teaching experiment, there was no significant difference among their writing skills.

### Mid-test scores

After the one-month teaching experiment, the mid-test and pre-test scores of the experimental class were analyzed, and at the same time, the mid-test and pre-test scores of the control class were also compared.

Table 3

*The pre-test and mid test score of test group comparison char*

	Number	Mean	Std.Deviation	P
Pro test	26	11.4230	3.3157	0.0571
Mid test	26	11.9615	3.4010	

From the table, it can be seen that, whether for the experimental class or the test class, the mid-test scores showed no significant changes. The average scores rose from 11.4230 to 11.9615 and from 11.4807 to 11.7500, respectively.

Table 4  
*The Pre-test and Mid Test Score of Control Group Comparison Char*

	Number	Mean	Std.Deviation	P
Pro test	26	11.4807	3.1838	0.0698
Mid test	26	11.7500	3.0861	

However, from the standard variances, it can be seen that the standard variance of the experimental class rose from 3.3157 to 3.4010, while the standard variance of the control class rose from 3.1838 to 3.0861, indicating that the gaps in the scores of the experimental class were gradually being widened.

### Post-test scores

After a semester of the teaching experiment, the post-test was carried out on the experimental class and the control class. The post-test scores were compared with the mid-test ones, as shown in Table 5 and Table 6

Table 5  
*The Post Test and Mid Test Score of Test Group Comparison Char*

	Number	Mean	Std.Deviation	P
Mid test	26	11.9615	3.4010	0.010
Post test	26	13.1923	3.7938	

Table 6  
*The Post Test and Mid Test Score of Control Group Comparison Char*

	Number	Mean	Std.Deviation	P
Mid test	26	11.7500	3.0861	0.4991
Post test	26	12.961	3.0179	

According to the data in the table, the average score of the experimental class in the mid-test was 11.9615, while the average post-test score was 13.1923, and  $P=0.01<0.05$ , indicating that the scores of the students in the experimental class improved significantly. While in the control class, the students' average score only slightly improved. Through the standard variance analysis, it is found that the standard variance of the experimental class was further reduced and gaps between the scores of different students continued to expand, while the scores of the students in the control class became more concentrated.

### Questionnaire survey

After a semester of experimental study, this paper randomly selected 20 students in the experimental class for a feedback survey, and the survey results are shown in Table 7.

Table 7  
*The Survey Results*

Survey points	Very helpful (Attitude)	Normal (Attitude)
The overall impression	18	2
The usage of process writing method	17	3
Impact on the level of English Writing	20	0

18 students have a good overall impression about the data-driven English writing teaching method and believe that it is very helpful to improving their English writing skills. 17 students have given high ratings on the process writing approach. And all students believe that, after receiving data-driven English writing teaching for a semester, their English writing skills have been improved.

After a semester of data-driven teaching in combination with the process writing approach, the post-test scores of the students in the experimental class are much higher than their pre-test scores and also higher than the post-test scores of those in the control class. Therefore, both the experimental analysis and student questionnaire survey results have proved the effectiveness of data-driven learning in improving college students' English writing skills.

## Conclusions

This paper first elaborates the theoretical basis for data-driven learning, and describes the application of corpus in the English writing learning process with an example. By properly setting up a teaching experiment and designing the teaching content, this paper completes the research on the effectiveness of data-driven learning in College English writing teaching. The conclusions and significance of this paper are as follows:

(1) Data-driven learning takes corpus as the main tool, which enriches the English teaching and learning methods and also expands college students' learning platforms, and allows them to better utilize the contextual resources of English-speaking countries.

(2) Data-driven learning can play a positive role in improving college students' English writing skills and the effects are remarkable.

(3) Data-driven learning has great effects on improving on students' self-learning abilities and learning initiatives.

## References

- Aleahmad, T., Aleven, V., & Kraut, R. (2009). Creating a corpus of targeted learning resources with a web-based open authoring tool. *IEEE Transactions on Learning Technologies*, 2(1), 3-9. <https://dx.doi.org/10.1109/TLT.2009.8>
- Cazorla, M. J. V. (2001). The teaching of modern languages in gran Canaria in the XIXth century, *Journal of Anatomy*, 123(2), 467-86.
- Espejo, R. J., Svalgaard, M., & Dyer, S. D. (2006). Characterizing fiber Bragg grating index profiles to improve the writing process, *IEEE Photonics Technology Letters*, 18(21), 2242-2244. <https://dx.doi.org/10.1109/LPT.2006.884740>
- He, H. (2016). A computer-aided analysis on word form errors in college English writing — A corpus-based study. *International Journal of Emerging Technologies in Learning*, 11(3), 4.
- Inanoglu, Z., & Young, S. (2009). Data-driven emotion conversion in spoken English. *Speech Communication*, 51(3), 268-283. <https://dx.doi.org/10.1016/j.specom.2008.09.006>
- Lughofer, E., & Kindermann, S. (2010). Sparsefis: data-driven learning of fuzzy systems with sparsity constraints, *IEEE Transactions on Fuzzy Systems*, 18(2), 396-411. <https://dx.doi.org/10.1109/TFUZZ.2010.2042960>

- Modhish, A. S. (2005). Teaching English writing skills to the speakers of Arabic pursuing a bachelor's degree in education: a study in EFL. *Experimental Mechanics*, 31(2), 93-97.
- Narita, M., Kurokawa, K., & Utsuro, T. (2003). Case study on the development of a computer-based support tool for assisting Japanese software engineers with their English writing needs. *IEEE Transactions on Professional Communication*, 46(3), 194-209. <https://dx.doi.org/10.1109/TPC.2003.816793>
- Peng, J. Y., Aston, J. A., Gunn, R. N., Liou, C. Y., & Ashburner, J. (2008). Dynamic positron emission tomography data-driven analysis using sparse bayesian learning. *IEEE Transactions on Medical Imaging*, 27(9), 1356-1369. <https://dx.doi.org/10.1109/TMI.2008.922185>
- Resnik, P., Smith, N. A. (2002). The web as a parallel corpus, *Computational Linguistics*, 29(3), 349-380. <https://doi.org/10.1162/089120103322711578>
- Solomatine, D. P., Maskey, M., & Shrestha, D. L. (2010). Instance-based learning compared to other data-driven methods in hydrological forecasting. *Hydrological Processes*, 22(2), 275-287. <https://doi.org/10.1002/hyp.6592>
- Talvensaari, T. (2008). Comparable corpora in cross-language information retrieval. *ACM Transactions on Information Systems*, 25(1), 79-82.
- Talvensaari, T., Laurikkala, J., Juhola, M., & Keskustalo, H. (2007). Creating and exploiting a comparable corpus in cross-language information retrieval, *ACM Transactions on Information Systems*, 25(1), 4. <https://doi.org/10.1145/1198296.1198300>
- Tong, L. C. (1989). A data-driven control strategy for grammar writing systems, *Machine Translation*, 4(4), 281-297.
- Turney, P. D., & Littman, M. L. (2005). Corpus-based learning of analogies and semantic relations. *Machine Learning*, 60(1-3), 251-278. <https://doi.org/10.1007/s10994-005-0913-1>