

# Level of Vocabulary as a Score Factor on the SAT Essay

B. YANG & W.K. He & S. Li

*School of Education, HuBei University, Wuhan, 430062, China*

T. T. GOH

*School of Information Management, Victoria University of Wellington, Wellington, 6011, New Zealand*

**ABSTRACT:** Writing requirements for SAT exam become stricter than the past after reformation. However, as a fundamental part of writing, vocabulary cannot be ignored. In this paper, in order to analyze the level of Chinese candidates' application of SAT vocabulary, we use MSSQL full-text search capabilities to retrieve probability of Chinese SAT essay which contains the 1000 Most Common SAT Words.

**KEYWORD:** SAT ESSAY; the application level of vocabulary; full-text search

## 1 INTRODUCTION

Vocabulary as a fundamental part of English writing, its application is an important scoring elements of the SAT writing automatic scoring and human raters [1]. And vocabulary, as one of the most important major factors of grading, can determine the quality of an article [2]. In traditional English SAT essay scoring, human raters think that longer essays usually score higher, and the length of the writing can reflect the writing skills of the candidates, and also the mastery of the vocabulary [3]. However, these ignore the spelling mistakes in the composition and the application level of vocabulary. In 2005, SAT essay score standard had gone through a series of great changes. The SAT essays are scored holistically, and a number of aspects are taken into account, including: the development of a point of view, the logical presentation of ideas, clear reasoning, sustained focus, appropriate choices of evidence, skillful coherence, effective organization, and precise use of language. After these reforms, the SAT writing scoring requirements become stricter than ever. Therefore, the examinee must be much more careful from the composition of the elements to prepare multiple score, application and vocabulary level [4].

For essay scoring introducing from abroad, the vocabulary usage has been regarded as one of significant criteria, which is closely related to composition score. From the point of view of Burstein and Martin Chodorow's experiment, the application level of automatic scoring system E - rater in vocabulary as a rating factor of composition score, high scores of the composition part usually behave better than low scores of the part in the

aspect of vocabulary usage, and human raters who rank selected sentence from a composition is more fastidious than a poor vocabulary usage composition score [1].

Nation points out that Foreign/Second Language learners in English (English Foreign/Second Language learners, EFL/ESL), and they hope that the basis of understanding and effective application of lexical information, because vocabulary usage will affect their communication and presentation to some extent [5]. Nassaji pointed out that people can use a kind of widely-used measures - vocabulary level test, with different level of word frequency, range from high frequency (2000) to the low-frequency words (10000) to the word matching test, and evaluation of the application level of vocabulary in the article [6]. Milton study, we can get effective and stable result through the matching tests [7].

Based on Nassaji's research methods and conclusions, this study on the SAT website (<http://sat.collegeboard.org/>) to acquire the 1000 most common SAT vocabulary, to match the SAT test China examinee composition of vocabulary application situation and analysis of Chinese students SAT composition of vocabulary application level.

## 2 RESEARCH DESIGN

### 2.1 *Research Object*

The object of study is students coming from several Chinese cities (Beijing, Shanghai, Taiyuan, Fuzhou, Guangzhou, Shenzhen, etc.) who are preparing for or interested in taking the SAT. Extracting the students' essays in the use of Duke education SAT simulation

test platform for composition training and simulation test which including essay exams, the number reached 600, each essay is scored by Duke education teacher in strict accordance with SAT essay score standard .Two trained essay raters holistically score each essay on a scale of 0 to 6, where 6 represents an outstanding essay demonstrating clear and consistent mastery, and 1 represents an essay that is fundamentally lacking or demonstrating minimal or no mastery.

SAT essay scoring standard evaluates score from writing 6 factors, including: arguments, organization, sentence, vocabulary application level, grammar, usage, overall performance. And each factor is divided into six levels, among them, the vocabulary application level apply to six grade levels: "6" represents a very mature application of words, "5" indicates abundant words, "4" represents the examination requirements enough words, "3" indicates inappropriate or wrong words, "2" very limited number of words, "1" indicates a limited vocabulary and a serious error.

## 2.2 Data source – the SAT essay

The 1000 most common SAT words refer to frequently appear in the SAT exam. SAT test index is in continuous development, and the most common etymology in the SAT is also in constant development, of which the purpose is still clear: it is to test the universe level who have reserves and generally have academic study of vocabulary reserves. Studies have repeatedly shown that advanced vocabulary is the most accurate indicator of academic success potential [8].

The study cites 1000 commonly used words come from the official website of the United States the SAT, and SAT agencies through statistical examination of numerous candidates come with a certain authority. This study is based on this vocabulary to detect the vocabulary level of Chinese students, whether to achieve the SAT test requirements.

## 2.3 Tool and Method

In this study, MSSQL 2008's full-text search function is used to query the data when T-SQL statement is a SELECT statement, but when you set the query conditions and the query SELECT statement setting somewhat different. In the T-SQL language, you can set conditions for full-text search query in the WHERE clause of the SELECT statement, you can also set the query in the FROM clause.

Full text search function of CONTAINS predicates can be simple word, derived words, prefixes, weighted words and adjacent words used in the data table five kinds

of search data. According to the actual situation of this study, consider the word have many in the composition of words in state select data search with the derivation of the word form.

CONTAINS syntax is as follows:

CONTAINS ({column\_name | (column\_list) | \*}, '< contains\_search\_condition >' [,LANGUAGE language\_term])

Among them, column\_name indicates that the field name; column\_list represents a list of field names; "\*" represents all fields; Contains\_search\_condition shows the search content of CONTAINS statements; LANGUAGE language\_term says with the settings for the query language.

Generation\_term derived word search as a way to search the Contains\_search\_condition, expressed as follows:

< generation\_term>::=FORMSOF ({ INFLECTIONAL|THESAURUS },< simple\_term >[,...n ] )

Among them, INFLECTIONAL is used to specify the use of language dependent stemmer. A Stemmer according to judge the stem of each specific language. THESAURUS: specifies the thesaurus.

With the help of the search mode search SAT 1000 common vocabulary in Chinese students' SAT writing in the frequency of occurrence, comparative analysis of other composition, the final analysis that the frequency and vocabulary scores, writing the final score whether there exists a relationship.

## 3 DATA PROCESSING

First of all, the thesis content, grading and 1000 common used SAT words imported into the SQL SERVER 2008 database, the database to build a composition table and a key word table.

Secondly, establishing a database a full-text index. To the thesis build a full-text index table, the following shows how to create a full-text index of this research, in this study, the essay SATTest database content listed in the table to establish a full-text index, described as follows:

Step1. use SATTest.

Step2. execute sp\_fulltext\_database 'enable'.

Step3. ft\_SATTe, execute sp\_fulltext\_catalog 'ft\_SATTest', 'create'.

Step4. execute sp\_fulltext\_table 'essay', 'create', 'ft\_SATTest', 'pk\_id'.

Step5. execute sp\_fulltext\_column 'essay', 'content', 'add'.

Step6. execute sp\_fulltext\_table 'essay', 'activate'.

Step7. execute sp\_fulltext\_catalog 'ft\_SATTest', 'start\_full'.

So far, the full-text index is set up.

Finally, according to the above Settings, the full text search algorithm was designed on the basis of function, achieve a search in the composition of the number of occurrences of SAT 1000 commonly used words. The key algorithm is as follows:

```

set @Rows = (select COUNT(*) from
keywords )
set @Row = 1
while (@Row <=@Rows)
begin
select @keyword = keywords from
keywords where keywords.id = @Row
select * from essay where
CONTAINS(content, @keyword)
set @Row = @Row + 1
end

```

#### 4 ANALYSIS

According to the above data processing, it is concluded that in 600 Chinese students composition contains SAT1000 commonly used word (hereinafter referred to as the "key words") number of composition, calculating the average score level vocabulary and writing average score based on Duke education teacher ratings, specific results in the following table 1:

Table1 Processing Result

	Num (N)	Freq. (F)	Voc. level (V)	Aver. score (S)
<i>Contains the key word composition (W1)</i>	57	9.5%	6.0	6.28
<i>Don't contain the key word composition (W2)</i>	543	90.5%	5.0	4.55

Compared to the data in table 1, it can be seen that contains a keyword composition (W1) probability is only 9.5%, not containing key words composition (W2), the probability of 90.5%, F(W1) is far less than F (W2),and it shows that the quantity of the key words Chinese students used in the composition is very limited, and in the aspect of vocabulary level average score and the score of composition, V (W1) greater than V (W2), S (W1) is greater than the S (W2),proves that there is a positive correlation between vocabulary level score and writing score, also reflects the higher of the vocabulary score that student master and usage , the higher of its composition score.

After the above illustration, from the following analysis is applied to the Chinese students SAT1000 prepare for common vocabulary less probability of reasons:

First of all, for English as a second language learning students, compared with the English district students in English writing has some inherent disadvantages. Burstein and Martin Chodorow through the experimental comparison of different regional students (English and non-English district district) SAT writing found that writing vocabulary of US-born English students much richer than nonnative English district students (such as Chinese, Africa, Spain)[1]. Comparing with Chinese students and English students, there is congenital disadvantage.

Secondly, EFL / ESL for Chinese high school students' English language skills (proficiency) level is inadequate, the lack of an effective presentation skills and the application of appropriate vocabulary when writing in English. Studies have shown that various levels of English proficiency for EFL / ESL learners, higher levels of student proficiency, written essay are performing better in coherence, vocabulary application, presentation skills, the score, the higher its composition[9]. That is, the Chinese students during the SAT writing, language is not authentic, vocabulary application level is not high, but is much more difficult to obtain high scores.

Finally, the research object is China senior high school students, demands to master the vocabulary not meet SAT.SAT1000 commonly used vocabulary is USA ETS (test center) on the basis of previous SAT test statistics are summed up, facing the global SAT candidate. Each examinee should master the application; and through the statistical analysis of SAT writing application of key word probability Chinese student is too low, said Ming. Chinese students in vocabulary level could not reach the SAT exam requirements. So the students also need to work harder, to master and skilled application, in order to improve the composition of the vocabulary level the score.

#### 5 CONCLUSION

The research samples of this study is selected from Beijing, Shanghai, Taiyuan, Fuzhou, Guangzhou, Shenzhen and so on many city through the SAT test platform of Duke education, and it also selected 600 composition which has some certain representativeness. However, the sample amount is not enough which may lead to the error of results of the study, and next, the study will collect as more and more extensive, more representative of the SAT writing.

This study relies on the SAT Duke education platform, next it will study vocabulary level by

analyzing the composition of the candidates, the corresponding push adaptive learning vocabulary and vocabulary learning program developed in later writing, tracking and monitoring of candidates writing words on the application, and then make the appropriate analysis push vocabulary learning, so the cycle continues to promote the improvement of vocabulary application level candidates to gain advantage in the fierce SAT exam.

Vocabulary usage, vocabulary standards as the basis for writing the most important part of teaching and learning should not be ignored [10]. Moreover, the current global students who took the SAT (United States) is becoming much more and the competition will also be much more intense than ever. Chinese students who want to achieve good results in the SAT exam must pay more attention to the most basic SAT vocabulary.

## REFERENCES

[1] Jill Burstein, Martin Chodorow, Automated Essay Scoring for Nonnative English Speakers.

- [2] Winerip, M. (2005, May 4). SAT essay test rewards length and ignores errors. *The New York Times*. Retrieved from <http://www.nytimes.com>.
- [3] Kobrin, J. K., Deng, H., & Shaw, E. J. (2007). Does quantity equal quality? The relationship between length of response and scores on the SAT essay. *The Journal of Applied Testing Technology*, 8 (1).
- [4] Walters, J., Wolf, Y., 1996. Language awareness in non-native writers: metalinguistic judgments of need for revision. *Language Awareness* 5 (1), 3- 25.
- [5] Nation, I.S.P., 2001. *Learning Vocabulary in Another Language*. Cambridge University Press, Cambridge.
- [6] Nassaji, H., 2004. The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Canadian Modern Language Review* 61, 107e134.
- [7] Milton, J., 2009. *Measuring Second Language Vocabulary Acquisition*. Multilingual Matters, Bristol, England.
- [8] [http://nthpstudentfacilitator.com/wp-content/uploads/2014/04/sat\\_vocab.pdf](http://nthpstudentfacilitator.com/wp-content/uploads/2014/04/sat_vocab.pdf), The SAT's Top 1000 Vocabulary Words.
- [9] Wenxing Yang & Ying Sun, 2011. The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education*.
- [10] Alemi, M. & Tayebi, A. (2011). The influence of incidental and intentional vocabulary acquisition and vocabulary strategy use on learning L2 vocabularies. *Journal of Language Teaching and Research*, 2(1).