EFFECTIVENESS OF AUTOMATED CHINESE SENTENCE SCORING WITH LATENT SEMANTIC ANALYSIS

Chen-Huei Liao
Department of Special Education
National Taichung University of Education, Taiwan
chenhueiliao@gmail.com

Bor-Chen Kuo
Graduate Institute of Educational Measurement and Statistics
National Taichung University of Education, Taiwan
kbc@mail.ntcu.edu.tw

Kai-Chih Pai
Graduate Institute of Educational Measurement and Statistics
National Taichung University of Education, Taiwan
minbai0926@gmail.com

ABSTRACT

Automated scoring by means of Latent Semantic Analysis (LSA) has been introduced lately to improve the traditional human scoring system. The purposes of the present study were to develop a LSA-based assessment system to evaluate children's Chinese sentence construction skills and to examine the effectiveness of LSA-based automated scoring function by comparing it with traditional human scoring. Twenty-seven fourth graders and thirty-one six graders were assessed on single-character sentence making test (subtest 1) and two-character words sentence making test (subtest 2). The outcomes of LSA-based automated scoring methods in three Chinese semantic spaces generated from three type weighting functions were compared to the traditional human scoring. The results showed that LSA-based automated scoring in three different Chinese semantic spaces and traditional human scoring were highly correlated in single-character sentence making test and moderately correlated in two-character words sentence making test. The Chinese semantic space generated from Log-IDF outperformed the other two types of weighting function in the present study.

INTRODUCTION

Writing skills are important for children's overall attainment. It is probably one of the few skills we learned in school that will be used often later in life. Writing is an essential element of children's education which has an impact on the progress of children achievement across the whole curriculum. Writing is also a means of communication; it allows children to participate actively in learning by sharing ideas, experience, thoughts, and feelings (Huang, Liu, & Hsiao, 2008). Effective writing, which requires writing with clarity, coherence, organization, and accurate grammar, is difficult to achieve, since it involves complex physical and mental processes. One important aspect that is fundamental in learning to write is constructing complete and grammatically correct sentences (Chik, Ho, Yeung, Wong, Chan, Chung, & Lo, 2010; Chik, Ho, Yenng, Chan, & Luan, 2011; Saddler, 2005).

Sentence construction can be as difficult a skill to assess as it is to learn. Reliable assessment requires a set of well-developed criteria and a significant amount of time devoted to the scoring procedure. In the present study, an automated scoring system with Latent Semantic Analysis (LSA) was developed to assess children's Chinese sentence construction skills. The system was designed as a pedagogical tool to provide instant computer-generated scores for sentence construction and to reduce the heavy load in the scoring process.

Latent Semantic Analysis (LSA) is a theory and method for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text (Landauer & Dumais, 1997; Landauer, McNamara, Dennis, & Kintsch, 2007). It is closely related to neural net models, but is based on singular value decomposition (SVD) and LSA used singular value decomposition to condense a large corpus of texts to 100-500 dimensions (Landauer, Foltz, and Laham, 1998; Landauer et al., 2007). The applications of LSA in educational settings were found in few studies. For example, Millis, Magliano, Wiemer-Hastings, Todaro, and McNamara (2007) assessed reading comprehension skills with LSA and found that LSA predicted reading comprehension skills and identified readers overall reading strategies. LSA was also involved in developing computer tutors, which provide instant feedbacks and teach conceptual knowledge to learners in Newtonian physics (VanLehn, Graesser, Jackson, Jordan, Olney, & Rosé, 2007) and computer literacy (Graesser, Lu, Jackson, Mitchell, Ventura, Olney, & Louwerse, 2004). Moreover, Graesser and his colleagues (Graesser, McNamara, Louwerse, & Cai, 2004; Graesser & McNamara, 2011; Graesser, McNamara, Kulikowich, 2011) developed a Coh-Metrix system with LSA to select appropriate texts for different levels of readers by providing multilevel analyses of text characteristics.

Past studies have shown that LSA has an enormous practical value in education; however, so far, LSA is not yet in the replacement of traditional human scorning. Therefore, the present study aimed at developing an automated scoring system of Chinese sentence construction skills with LSA by comparing the effects of three semantic spaces that were established by different types of weighting function (Log-Entropy, Log-IDF, TF-IDF). Few studies discussed the utility of applying different types of weighting function in LSA and found that Log-Entropy gave better results than the other proposed methods (Dumais, 1991; Lintean, Moldovan, Rus, & McNamara, 2010; Nakov, Popova, & Mateev 2001). Thus, generally in application, Log-Entropy was used to develop the semantic space of LSA (Chen, Wang, & Ko, 2009; Quesada, 2006). Nevertheless, empirical evidence supporting the application of various types of weighting function in LSA is still scarce. In this study, three semantic spaces were developed by adopting three types of weighting function and the performance was examined. Finally, the effectiveness of LSA-based automated scoring system was examined by comparing the correlations between human scoring and LSA-based automated scoring.

Latent Semantic Analysis

To make use of LSA, establishing a semantic space to represent the type-by-document matrix in a given corpus in which each row stands for unique type and each column stands for a document is required. Each element of the type-by-document matrix contains the frequency with which the type of its row appeared in the passage denoted by its column. The type-by-document matrix is often transformed to weight them by their estimated importance in order to better mimic human comprehension process (Landauer et al., 1998; Martin & Berry, 2007; He, Hui, & Quan, 2009; Olmos, León, Escudero, & Jorge-Botana, 2011).

Next, SVD (singular value decomposition) and dimension reduction to the type-by-document matrix is applied. SVD is the method used by LSA to decompose the type-by-document input matrix **A**. The SVD for $m \times n$ type-by document input matrix **A** with the rank of **A**=r is defined as follows:

$$A = U \Sigma V^T \qquad \text{Equation 1}$$

Where U is an orthogonal matrix, V is an orthogonal matrix, and Σ is a diagonal matrix with the remaining matrix cells all zeros (Berry & Browne, 2005; Golub & van Loan, 1989). Dimension reduction is used to remove the extraneous information and variability in type and document vectors which referred to as “noise”. A pictorial representation of the SVD of input matrix **A** and the best rank-k approximation to **A** is shown in Figure 1 (Berry, Dumais, & O’Brien, 1995; Martin & Berry, 2007; Witter & Berry, 1998).

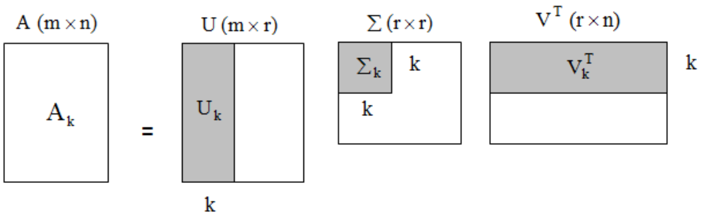


Figure 1. Diagram of the truncated SVD

After SVD and dimension reduction, A_k is the k-dimensional vector space which is called “semantic space”.

Objectives of the study

1. To develop a LSA-based assessment system to evaluate children’s Chinese sentence construction skills. To develop a LSA-based assessment system to assess sentence construction skills, single-character sentence construction test (subtest 1) and two-character words sentence construction test (subtest 2) were constructed by two instructors of language and literacy education department.

2. To examine the effectiveness of LSA-based automated scoring function by comparing it with traditional human scoring. To develop the automated scoring system, LSA was employed and the effectiveness of the automated scoring system was examined by the results obtained by human

raters and the system. In addition, the effects of three semantic spaces that were established by different types of weighting function (Log-Entropy, Log-IDF, TF-IDF) were also examined.

Research Questions

1. Does LSA-based automated scoring system score children's performance on sentence construction tests as well as human raters?

2. Does the Chinese semantic space generated from Log-Entropy outperform the Chinese semantic spaces generated from Log-IDF and TF-IDF?

METHOD

Participants

There was a total of 58 participants (27 fourth graders and 31 six graders) at Sin-Yi elementary school in Taichung, Taiwan. The mean age of the participants was 10.8 years (range 9.3 to 12.2, SD =1.03). None of the children was previously diagnosed with any emotional, behavioural or sensory difficulties.

Sentence Construction Tests

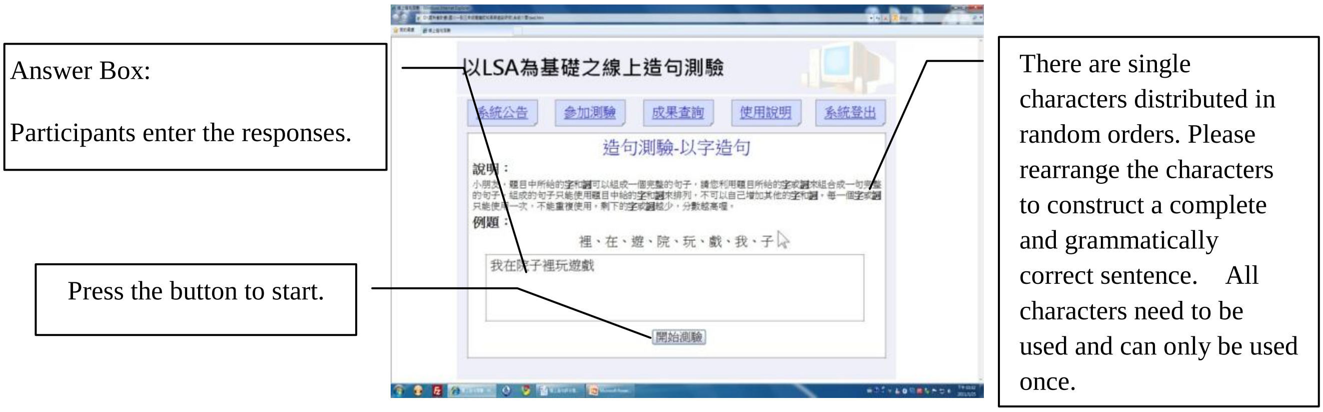
Sentence construction skills were assessed by two subtests: single-character sentence construction test (subtest 1) and two-character words sentence construction test (subtest 2). The subtests took approximately 40 minutes to finish. All the tests were computerized.

Single-character sentence construction test (subtest 1)

There were two practice trials and 10 test trials. In each trial, Chinese single characters were distributed in a raw in random order. Participants were asked to rearrange all the given characters to construct a complete and grammatically correct sentence (an item example is shown in Table 1). The number of characters in each test item ranged from 8 to 16 characters. The interface and instruction of single-character sentence construction is illustrated in Figure 2.

Table 1. An example of single-character sentence construction test

Item	裡、在、遊、院、玩、戲、我、子
Answer	我在院子裡玩遊戲 I play games in the yard



Answer Box:

Participants enter the responses.

Press the button to start.

There are single characters distributed in random orders. Please rearrange the characters to construct a complete and grammatically correct sentence. All characters need to be used and can only be used once.

以LSA為基礎之線上造句測驗

造句測驗-以字造句

說明：

小朋友，題目中所給的字和詞可以組成一個完整的句子，請您利用題目所給的字或詞來組合成一句完整的句子，組成的句子只能使用題目中給的字和詞來排列，不可以自己增加其他的字和詞，每一個字或詞只能使用一次，不能重複使用，剩下的字或詞越少，分數越高喔。

例題：

裡、在、遊、院、玩、戲、我、子

我在院子裡玩遊戲

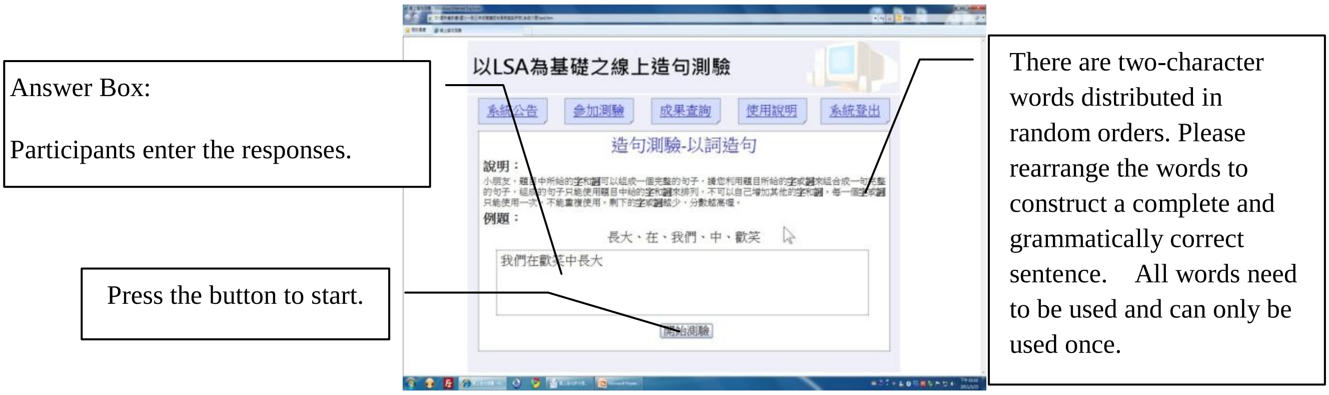
Figure 2. Interface of the single-character sentence construction test

Two-character words sentence construction test (subtest 2)

There were two practice trials and 10 test trials. In each trial, Chinese two-character words were distributed in a raw in random orders. Participants were asked to rearrange all the words provided to construct a complete and grammatically correct sentence (an item example is shown in Table 2). The number of words in each test item ranged from 5 to 8 words. The interface and instruction of two-character words sentence construction test is illustrated in Figure 3.

Table 2. An example of two-character words sentence construction test

Item	長大、在、我們、中、歡笑
Answer	我們在歡笑中長大 We grew up with laughter and joy.



造句測驗-以詞造句

說明：

小朋友，題目中所給的字和詞可以組成一個完整的句子，請您利用題目所給的字或詞來組合成一句完整的句子，組成的句子只能使用題目中給的字和詞來排列，不可以自己增加其他的字和詞，每一個字或詞只能使用一次，不能重複使用，剩下的字或詞越少，分數越高喔。

例題：

長大、在、我們、中、歡笑

我們在歡笑中長大

Answer Box:

Participants enter the responses.

Press the button to start.

There are two-character words distributed in random orders. Please rearrange the words to construct a complete and grammatically correct sentence. All words need to be used and can only be used once.

Figure 3. Interface of the two-character words sentence construction test

Human Scorning
Two Chinese literacy teachers played the role as human raters. The scores of test items were given based on the number of characters/words used and the grammatical correctness of the sentence. Taking the test item in Table 1 as an example, if a participant constructed a grammatically correct sentence with all the eight given characters, he /she would get a full score which is 8, for this particular item. However, if the participant only used six out of the eight words to construct a grammatically correct sentence, he/she would get 6 for the item.

Chinese corpus
The present study used Academia Sinica Balanced Corpus of Modern Chinese (3.1) from Academia Sinica in Taiwan to establish Chinese semantic spaces of LSA. The corpus contained 5 million words and 9227 documents.

Types of weighting function
A weighting function is generally applied to each nonzero (type frequency for type i in document j) element, a_{ij}, of the matrix **A** to improve retrieval performance (Berry & Browne, 2005; Dumais, 1991). LSA applies both a local and global weighting function to each nonzero element, a_{ij}, in order to increase or decrease the importance of types within documents (local) and across the entire document collection (global). So a_{ij} = local(i, j)*global(i), where local(i, j) is the local weighting for type i in document j, and global (i) is the type's global weighting (Dumais, 1991; Letsche & Berry, 1997). The study used three different types of weighting function: Log-Entropy, Log-IDF, TF-IDF, and the equations are as follow:

$$\begin{cases} L(i,j) = \log(tf_{ij}+1) \\ G(i) \ = 1 + \sum_{\mathrm{j}} \dfrac{p_{ij}\log_2(p_{ij})}{\log_2 n},\ p_{ij} = \dfrac{tf_{ij}}{gf_i} \end{cases} \qquad \text{Equation 2}$$

Equation 2 is the type weighting function of Log-Entropy, where $L(i, j)$ represents local weighting. tf_{ij} represents type frequency of type i in document j. $G(i)$ represents global weighting, and gf_i represents the total number of times that type i appears in the entire collection of n documents.

$$\begin{cases} L(i,j) = \log(tf_{ij}+1) \\ G(i) = \log(m/df(i)) \end{cases} \qquad \text{Equation 3}$$

Equation 3 is the type weighting function of Log-IDF, where m is the total number of documents and $df(i)$ is used the document frequency.

$$\begin{cases} L(i,j) = \dfrac{n_{i,j}}{\sum_{k}^{N} n_{k,j}} \\ G(i) = \log(m/df(i)) \end{cases} \qquad \text{Equation 4}$$

Equation 4 is the type weighting function of TF-IDF, where $n_{i,j}$ is the number of times the type i occurs in the given document j , $n_{k,j}$ is the total number of types in the document.

LSA-based automated scoring

The ability to add new types and documents to reduce rank type-document vector space is important because the original information in the document collection often needs to be augmented for different contextual or conceptual usages (Martin & Berry, 2007). In the present study, a simple method of handling the addition of sentences was used by applying the fold-in procedure (Equation 5). Here, following the fold-in procedure, a new sentence folds into the existing k-dimensional vector space (Berry, Dumais, & O'Brien, 1995). As well, based on the existing type-document vector space, the fold-in procedure was applied to measure the similarity between the best answer and each participant's answer (Equation 6). A best answer was defined as the response that matches the correct answer in the system. :

$$d_{new} = d^T U_k \Sigma_k^{-1} \qquad \text{Equation 5}$$

In Equation 5, the vector *d*, represents the best answer or participants' answer, which contains zero and nonzero elements; where the nonzero elements correspond to the type frequencies contained in the sentence adjusted by term weighting function.

$$sim(S_1, S_2) = \frac{d_1 d_2^T}{\|d_1\|\|d_2\|} \qquad \text{Equation 6}$$

In Equation 6, the similarity is computed as the cosine of the vector representation of the sentences. d_1 represents the vector representation of the best answer, S_1, represents the best answer, and d_2 represents the vector representation of the participant's answer, and $S_{2,}$ represents the participant answer.
Finally, LSA-based automated scoring equation (Equation 7) is presented as follows:

$$score_{item} = sim(S_1, S_2) * s_{item} \qquad \text{Equation 7}$$

s_{item} represents the maximum score in each item, $sim(S_{1,}\ S_2)$ represents the semantic similarity between the correct answer and the participant's answer. And $score_{item}$ represents the participant's sentence construction score in each item.

RESULTS

Pearson correlations between human scoring and LSA-based automated scoring were calculated to examine the effectiveness of LSA-based automated scoring. The study used three types of weighting function and the results were presented in Table. 3, Table. 4, and Table. 5.

Table 3. Correlations between human scoring and LSA-based automated scoring (Log-Entropy)

Variable	1	2	3	4
1 human scoring (subtest 1)	1	—	—	—
2 LSA-based automated scoring (subtest 1)	0.912**	1	—	—
3 human scoring (subset 2)	0.710**	0.611**	1	—
4 LSA-based automated scoring (subtest 2)	0.511**	0.522**	0.531**	1

*p<0.05; **p<0.01

Table 4. Correlations between human scoring and LSA-based automated scoring (Log-IDF)

Variable	1	2	3	4
1 human scoring (subtest 1)	1	—	—	—
2 LSA-based automated scoring (subtest 1)	0.916**	1	—	—
3 human scoring (subset 2)	0.710**	0.617**	1	—
4 LSA-based automated scoring (subtest 2)	0.508**	0.524**	0.543**	1

*p<0.05; **p<0.01

Table 5. Correlations between human scoring and LSA-based automated scoring (TF-IDF)

Variable	1	2	3	4
1 human scoring (subtest 1)	1	—	—	—
2 LSA-based automated scoring (subtest 1)	0.901**	1	—	—
3 human scoring (subest2)	0.710**	0.594**	1	—
4 LSA-based automated scoring (subtest 2)	0.489**	0.487**	0.467**	1

*p<0.05; **p<0.01

The results showed in subtest 1, human scoring and LSA-based automated scoring were strongly correlated (*rs* = 0.912, 0.916, 0.901). In subtest 2, however, human scoring and LSA-based automated scoring were moderately correlated (*rs* = 0.531, 0.543, 0.467). Moreover, the relations between LSA-based automated scoring and human scoring were more consistent in subtest 1 than in subtest 2. Moreover, contrary to what were found in previous studies (e.g., Dumais, 1991; Lintean, Moldovan, Rus, & McNamara, 2010), our results showed that the automated scorning system established by semantic space of Log-IDF worked slightly better than the two other methods (Log-Entropy and TF-IDF) in subtest 1 (*rs* = 0.916, 0.912, 0.901) and subtest 2 (*rs* = 0.543, 0.531, 0.467). The outcomes of the three types of weighting function showed that the Chinese semantic space generated from Log-IDF outperformed the other two types of weighting function (Log-Entropy and TF-IDF).

CONCLUSION

The present study developed LSA-based assessment system and examined the effectiveness of LSA-based automated scoring function by comparing it with traditional human scoring. The results showed that, in subtest 1(single-character sentence construction test), LSA-based automated scoring and human scoring were highly correlated in three types of weighting function, which implies that LSA-based automated scoring was comparable to human scoring. In subtest 2 (two-character words sentence construction test), LSA-based automated scoring and human scoring were only moderately correlated, which implies that human raters and LSA did not score children's sentence construction skills equivalently. It was interesting to discover that LSA-based automated scoring system acted similar to human raters in single-character sentence construction test (subtest 1) but less well to two-character words sentence construction test (subtest 2). LSA automated scoring system rated children's answers by comparing them with the pre-set best answers. However, one of the well-known limitations of LSA is that it made no use of word order, syntactic relations or logic, and morphology (Landauer et al., 1998). In subtest 2 (two-character words sentence construction test), the rearrangement of the two-character words produced high similarities between grammatically incorrect sentences and the best answers provided by the automated scoring system. In Chinese, each character is a morpheme, and morphemes are combined into words. Most of Chinese words involve multiple morphemes, for example, 天空 *sky*, 美麗 *beautiful*, 我們 *we*, are two-character (morpheme) words. Therefore, the ability to manipulate and to be aware of morphemes (characters) is important for Chinese literacy acquisition. In LSA Chinese scoring system, when a sentence (or a row of characters) is given, the system automatically segments the row of characters into words that match the corpus (e.g. 藍藍的/天空/很/美麗, *The blue sky is beautiful*). However, in subtest 2, the "two-character words" were provided in the test items and therefore, as long as the participant used all the given two-character words, the answers would automatically match the " pre-set answers" in the system. Hence, even the sentence was grammatically and syntactically incorrect, a high score would still be given by the system due to the great resemblance between the responses and the pre-set best answers. Therefore, the equivalency of scoring was not met between human raters and the system. On contrary, in subtest 1, only single characters were given, thus, the participants were required to recognize every character, to combine all the given characters into meaningful words, and to construct grammatically and syntactically correct sentences with these words. The skills and behaviors require in subtest 1 actually bear a resemblance to the actually writing activity. Consequently, human raters and the system scored children's performance on single-character sentence construction test similarly. Moreover, in subtest 1 (single-character sentence construction test), the present automated scoring system captured both children’s morphological processing skills and sentence construction skills. In addition, the results of the present study did not support that Log-Entropy is more appropriate in developing the Chinese LSA-based automated scoring. One possible explanation is that previous studies were conducted in English. The characteristics of Chinese may require a different method of weighting function to reflect the nature of the language. In conclusion, LSA-based automated scoring system is effective in assessing children’s sentence construction skills and Chinese semantic space generated from Log-IDF is better compare to the other two types of weighting function for the automated scoring mechanism.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of this study by the National Science Council of Taiwan, under the Grant No. NSC-100-2420-H-142-001-MY3. The study also thanks The Association for Computational Linguistics and Chinese Language Processing (ACLCLP) and The CKIP (Chinese Knowledge and Information Processing).

REFERENCES

Berry, M. W., & Browne, M. (2005). Understanding search engines: Mathematical modeling and text retrieval. *Philadelphia: SIAM*, 2.

Berry, M. W., Dumais, S. T., & O’Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. *SIAM Review*, *37*, 573–595.

Chen, M. L., Wang, H. C., & Ko, H. W. (2009). The construction and validation of Chinese semantic space by using latent semantic analysis. *Chinese Journal of Psychology*, *51*, 415-435.

Chik, P. P., Ho, C. S., Yeung, P., Chan, D. W., Chung, K. K., & Luan, H. (2011). Syntactic skills in sentence reading comprehension among Chinese elementary school children. *Reading and Writing: An Interdisciplinary Journal*, *4*, 1–22.

Chik P. P., Ho, C. S., Yeung, P. S., Wong, Y. K., Chan, D. W., Chung, K. K., & Lo, L. Y. (2010). Contribution of discourse and morphosyntax skills to reading comprehension in Chinese dyslexic and typically developing children. *Annals of Dyslexia*. Retrieved from http://www.springerlink.com/content/w61239486x05p205/

Dumais, S. T. (1991). Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments, and Computers*, *23*, 229–236.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, *36*, 180–193.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*, 193-202.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*, 371-398.

Graesser, A.C., McNamara, D.S., & Kulikowich, J. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*, 223-234.

Golub, G., & Van Loan, C. F. (1989). *Matrix computations*. (2nd ed.), Johns Hopkins, Baltimore.

Huang, T. H., Liu, Y. C., & Hsiao, W. T. (2008). *Research on the influence of computer network supported cooperative learning on sentence construction skills of elementary school students*. Paper presented at the meeting of the 38th ASEE/IEEE Frontiers in Education Conference, Saratoga Spring, USA.

He, Y., Hui, S. C., Quan, T. T. (2009). Automatic summary assessment for intelligent tutoring systems. *Computers & Educat*ion, *53*, 890-899.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, *25*, 259-284.

Landauer, T. K., Laham, D., Rehder, B., & Schreiner, M. E., (1997). How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th annual meeting of the Cognitive Science Society* (pp. 412-417). Mawhwah, NJ: Erlbaum.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Lawrence Erlbaum.

Letsche, T. A., & Berry, M. W. (1997). Large-scale information retrieval with latent semantic indexing. *Information Sciences*, *100*, 105–137.

Lintean, M., Moldovan, C., Rus, V., & McNamara D. S. (2010). *The role of local and global weighting in assessing the semantic similarity of texts using Latent Semantic Analysis.* Proceedings of the 23rd International Florida Artificial Intelligence Research Society Conference. Daytona Beach, FL.

Martin, D.I., & Berry, M. W. (2007). Mathematical foundations behind Latent Semantic Analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*. (pp. 35-55). Mahwah, NJ: Lawrence Erlbaum Associates.

Millis, K. K., Magliano, J. P., Wiemer-Hastings, K., Todaro, S., & McNamara, D. S. (2007). Assessing and improving comprehension with Latent Semantic Analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*. (pp. 207-225). Mahwah, NJ: Lawrence Erlbaum Associates.

Nakov, P., Popova, A., and Mateev, P. (2001). Weight functions impact on LSA performance. *In Proceedings of the Euro Conference Recent Advances in Natural Language Processing*, Bulgaria, pp.187-193.

Olmos, R., León, J. A., Escudero, I., & Jorge-Botana, G. (2011). Using latent semantic analysis to grade brief summaries: some proposals. *International Journal of Continuing Engineering Education and Life-Long Learning*, *21*, 192-209

Quesada, J. (2007). Creating your wwn LSA spaces. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), Handbook of Latent Semantic Analysis. (pp. 71-85). Mahwah, NJ: Lawrence Erlbaum Associates.

Saddler, B. (2005). Sentence combining: A sentence-level writing intervention. *Reading Teacher*, *58*, 468-471.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, *31*, 3-62.

Witter, D., & Berry, M. W. (1998). Downdating the latent semantic indexing model for conceptual information retrieval. *The Computer Journal*, *41*, 589–601.