

## Effects of text structures on children's reading comprehension

Feng-Lan Kuo

Heien-Kun Chiang

*National Changhua University of Education, Taiwan*

### Reference data:

Kuo, F. L., & Chiang, H. K. (2009). Effects of text structures on children's reading comprehension. In A. M. Stoke (Ed.), *JALT2008 Conference Proceedings*. Tokyo: JALT.

Text types have long been found to affect the reading comprehension or information recall of adult or adolescent language learners (Alvermann & Boothby, 1982; Carrell, 1985; Law, 1995). Furthermore, recent studies have shown that L1 children often encounter difficulties in transition from reading fiction in the lower grades to reading nonfiction in the upper grades. Camp (2000) therefore suggested that pairing fiction and nonfiction books on the same topic, along with interactive class strategies, can encourage children to read for comprehension and enjoyment. However, only impressionistic positive results on five case studies have been reported. This study then aims to investigate the impact of text structures and language proficiency levels on reading comprehension of 5th graders in Taiwan. The subjects were 119 students of three reading proficiency levels. A compare/contrast expository text paired with its respective fiction text were used as the testing materials. Multiple-choice comprehension tests and written recalls were used to measure the learners' comprehension of the two types of texts. Though the pair books of the twin texts were judged by the publisher to be of equal difficulty level, results of this study showed that the fifth graders performed significantly better in comprehending the fiction text than the nonfiction text. Complied with the findings of previous studies conducted in L1 environment, the results indicated that the nonfiction text posed greater reading difficulty than the fiction text for the Chinese EFL learners. Results further revealed that text structure and language proficiency levels are factors affecting these learners' reading comprehension.

**R**ecent studies have shown that L1 children often encounter difficulties in transition from reading fiction in the lower grades to reading nonfiction in the upper grades (Cheuvront, 2002; Dymock, 2005). Elementary school children in both L1 and L2 environments typically were found to read more narrative texts (fiction books) than expository texts (nonfiction books). It was steadfastly maintained



that children prefer storybooks to other genres, such as informational books. Alvermann and Boothby (1982) surveyed thirty-eight 4th graders in a public school in a small city in Iowa on text differences between a narrative passage of 4th grade reading level and an expository passage of 5th grade reading level. The results showed that the narrative passage was identified as easier than the expository passage by 70% of the children. This indicated that in an L1 environment children in the 4th grade, the transition stage, often encountered difficulties in reading expository texts. The imbalance between narrative and expository experience might have led to reading difficulty of 4th graders in an L1 context. Camp (2000) therefore suggested that pairing fiction and nonfiction books (so called twin-texts) on the same topic along with interactive class strategies can encourage children to read for comprehension and enjoyment. However, only impressionistic positive results on five case studies have been reported. Furthermore, the difficulty levels of the twin-texts were unknown. Since English education in Taiwan was first implemented at the elementary school level in 2001 and elementary school children were mostly exposed to narrative texts, the difficulties of reading expository texts need to be identified and the appropriateness of using twin-texts at the elementary school level needs to be established. Thus, empirical studies on the effects of text structures on reading comprehension are called for. This study then aims to investigate the impact of text structures and language proficiency levels on reading comprehension of 5th graders in Taiwan in order to understand Chinese EFL young learners' reading difficulties. The following two research questions will be answered in this study.

1. Does text structure affect Chinese EFL 5th graders' reading comprehension?
2. Is there any relationship between text structure and learners' reading proficiency levels?

## Method

### Subjects

One hundred and nineteen participants from 4 intact classes of three reading proficiency levels participated in this study. The participants officially started receiving English classes at grade three. They received 40-minutes English instruction per week at third and fourth grade and 80-minutes English instruction per week at fifth grade. Before the experiment, the four groups were given an English reading proficiency test under two modified commercial English proficiency tests, STYLE (Level 3) and Cambridge YLE Starters (Level 1) and Movers (Level 2) to establish the homogeneity of the four intact classes and to establish their proficiency. The participants' scores on the proficiency test ranged from 23 to 98 points. Results of the ANOVA test indicated that there were no significant differences among the four groups on their homogeneity,  $F(3, 115) = 1.92, p = .13$ . Based on the reading scores, the subjects were further categorized into 3 proficiency groups, high proficiency group (HP), mid proficiency group (MP), and low proficiency group (LP).

### Testing materials

The reading materials used in this study were paired fiction and nonfiction texts, "Planet Star" and "Shapes." "Planet

Star,” a 12-page fiction book, was written by Julia Stanton and illustrated by Tania Konstant. “Shapes,” a 12-page nonfiction book, was written by Jane Keys. Both paired texts were taken from a basal reader series, *Twin Texts* readers (2002). *Twin Texts* readers are a series of paired books of fiction and nonfiction texts on related topics for emergent and young readers to make the transition from narrative to expository text. Among the 96 books, *Twin Texts* readers are divided into 6 sets and 24 instructional levels, with 2 pairs of fiction and nonfiction texts at each level and 16 books in each set. While the fiction texts are presented with clear story structures and vivid illustrations, the nonfiction texts are targeted to introduce new topics in an informational manner with factual pictures. The following three criteria were used for selecting the two books as the reading materials: relevance of topics, genre-oriented text structure, and level of difficulty. First and foremost, the two texts were selected because relevant topics were present as suggested by Camp (2000) who claimed that the paired texts should be on the same or related topics. Therefore, the two experimental texts were selected because they contained similar information on the topic of “Shapes.” Secondly, both texts were considered to have genre-oriented text structure. That is, the fiction text was composed with clear story elements as suggested by Schneider and Vis Dub’e (2005) while the nonfiction text was with a comparison/contrast text structure pattern (Moss, 2004). Finally, both texts were of the same reading level of *Twin Texts* (Reading Recovery Level 9, Guided Reading Level F) judged by the publisher to have the same level of difficulty.

### **Instruments**

Two test formats, a multiple-choice test and a written recall test were conducted separately. The instruments consists of a written recall of a fiction text (FT), a multiple-choice reading comprehension test of a fiction text; a written recall of a nonfiction text (NT) and a multiple-choice reading comprehension test of a nonfiction text. The subjects were asked to read the fiction book followed by the nonfiction book. They were first given the written recall test followed by the multiple-choice test to measure their reading comprehension of the books of two types of text structures.

Irwin and Mitchell’s (1983) procedure was applied to analyze the subjects’ written recall protocols. Irwin and Mitchell claimed that the scale was suitable for accessing students’ performance on both type of texts, fiction and nonfiction. They further suggested that students were allowed to use their own language to retell a given text. In this way, the subjects in this study were allowed to use Chinese (their mother language), English (the target language), or combination of the two languages. The Richness of Retelling Scale was adopted to assess the richness of students’ written recall protocol. Two graduate students served as the raters of the free recall for this study. In order to ensure the reliability of marking, 25% of the sample of 119 students (n=30) was randomly selected for an inter-rater reliability check using Spearman Rank Correlation. The inter-rater agreement between the two raters was .996 for scoring the fiction written recall test and .997 for nonfiction written recall test. Inter-rater reliability was .992 in total for both recall tests. After the inter-rater reliability had been established, the two raters discussed the

content over which there was disagreement and agreed on a rating in cases of discrepancy.

## Results and discussion

### *Text analysis of the two texts*

This study first provides a computer assisted text analysis of the two selected texts using word counts, average length per sentence, percentage of occurrences of passive sentences, the Flesch Reading Ease (Flesch, 1949), and the Flesch-Kincaid Grade Level to assess the readability of these two passages. Additionally, type-token ratio (TTR), lexical density, and vocabulary difficulty were used to compare the readability between the two texts. It was found that the fiction text "Planet Star" is 209 words long; the average length per sentence is 6.7 words; there is no passive sentence; the Flesch Reading Ease value is 98.7 (the higher the value, the easier the text is to read); the computed Flesch-Kincaid Grade Level is 1.1 (the score provides an estimate of the grade level required to understand the text). In contrast, a text analysis of the nonfiction text "Shapes" revealed that it is 116 words long; the average length per sentence is 5.2 words; there is no passive sentence; the Flesch Reading Ease value is 92; the computed Flesch-Kincaid Grade Level is 1.7.

TTR is often taken as an indicator of lexical diversity. Conventionally, it is calculated by dividing the number of different words (types) by total running words (tokens). Text analyses showed that the TTR for the fiction text "Planet Star" was 36 (0.36) while that for the nonfiction text "Shapes" was 39 (0.39). As found in Ying (2001), it seemed that the types that appeared in the first 100 running words

would be repeated in the subsequent text. In other words, fewer additional types would be added after the first 100 words. Therefore, if only the first 100 words of each text were calculated for lexical variety, the TTR for the fiction text was 55 (0.55) whereas the TTR for the nonfiction text was 43 (0.43). This suggested that the nonfiction text was harder to read since there were more different lexical items. The results further indicated that length of the texts affected the ratio. This provided support to the findings of Hess, Haug, and Landry (1989) and Ying (2001) that text length might largely affect TTR measurement.

Differing from TTR, lexical density, an indicator of the degree of literacy or orality of any text, measures the relationship between lexical items (content words) and grammatical items (function words). In general, written texts include a higher ratio of lexical items than spoken texts. Thus, written texts usually display a higher lexical density, as found in Ure's (1971) study showing that spoken texts typically had a lexical density under 40%, whereas written texts over 40%. In calculating lexical density, Halliday (1985) proposed two different ways to count lexical diversity: overall lexical density and weighted lexical density. Overall lexical density refers to the ratio of the number of lexical items to the total running words. On the other hand, weighted lexical density further differentiated high-frequency lexical items from low-frequency lexical items, with the former type given half the weight of the latter type. Then the total weight of the high- and low- lexical items was expressed as a proportion of the total running words. Though two different measures were proposed, Halliday (1985) had suggested that weighted lexical density

was more accurate and refined than overall lexical density because repetition of lexical items would reduce the effect of density. Results of prior studies have shown that lexical density was an effective means in measuring lexical diversity of texts of any length.

The results of present study showed that the fiction text had an overall lexical density of 59% and a weighted lexical density of 32.5%. Similarly, the nonfiction text had an overall lexical density of 60.5% and a weighted lexical density of 32.1%. This indicated that the two texts were of approximately equal lexical diversity.

Furthermore, vocabulary difficulty analysis (checking against Jeng's (2001) *CEEC English Word List* of 6,480 concept words classified into six levels of difficulty) showed that most of the vocabulary in the fiction text were high frequency words with 94% of them being Level 1 vocabulary whereas only 87% of the vocabulary of the nonfiction text were Level 1 vocabulary. This indicated that the vocabulary in the nonfiction text was more difficult than that in the fiction text.

In summary, text analyses showed that the two texts differed in word count, type-token ratio if text length was taken into consideration, and vocabulary difficulty. Though the Reading Recovery Level and the Guided Reading Level provided by the publisher as well as the Flesch Reading Ease and Flesch-Kincaid Grade Level judged the two texts to be of similar difficulty level, empirical evidence was needed to show whether language learners perform equally well in understanding the two texts. Results of this study are shown in the following two sections.

### Results of the multiple-choice tests

As shown in Table 1, the ANOVA analysis yielded a significant difference between the scores of the fiction multiple-choice test and those of the nonfiction multiple-choice test ( $F = 25.00, p < .05$ ). Results showed that the subjects performed significantly better in comprehending the fiction text than the nonfiction text in the multiple-choice test. Furthermore, the results did not reveal significant interaction effects between text structure and language proficiency.

**Table 1. Two-way ANOVA summary for text structure and proficiency in the multiple-choice test**

Source	DF	SS	MS	F	P
Text structure	1	8594.26	8594.26	25.00	.000*
Proficiency	2	17105.51	8552.76	24.88	.000*
Text structure x proficiency	2	858.08	429.04	1.25	.29

DF = degree of freedom; SS = sum of squares; MS = mean square

Taking subjects' proficiency level into account, the results also showed a significant difference among the three proficiency groups ( $F = 24.88, p < .05$ ). Table 2 shows the means and standard deviations of the scores for the multiple-choice test for the HP, MP, and LP learner groups in the fiction task and nonfiction task. In the FT multiple-choice test (ANOVA yielded significant statistics,  $F(3, 2) = 17.58, p < .001$ ), Scheffe multiple group post hoc comparisons revealed that the HP group scored significantly higher than

**Table 2. Means and standard deviations of the scores for the multiple-choice test for HP, MP, and LP learner groups in the Fiction Task and the Nonfiction Task**

Language group	N	FT structure		NT structure	
		M	SD	M	SD
High proficiency	40	5.55	2.87	3.90	2.57
Mid proficiency	39	3.54	1.76	2.31	1.36
Low proficiency	40	3.30	1.44	2.58	1.32
Total	119	4.13	2.11	2.93	1.96

Maximal score = 10; SD = standard deviation

the MP and LP groups. However, the latter two groups did not differ significantly from each other. Similarly, in the NT multiple-choice test (ANOVA yielded significant statistics,  $F(3, 2) = 8.47, p < .001$ ), post hoc comparisons showed that the HP group performed significantly better than the MP and LP groups. However, the latter two groups did not differ significantly from each other. There was no interaction effect between text structure and proficiency. Across different text structure, the HP group was found to score significantly higher in the FT,  $t = 4.76, p < .001$ . Likewise, the MP performed significantly better in the FT,  $t = 3.09, p = .004$ . In a similar vein, the LP group scored significantly higher in the FT,  $t = 2.66, p = .011$ .

### Results of the written recall tests

As shown in Table 3, in the free-recall test, the subjects performed equally poor in the FT and in the NT ( $F = 2.14$ ,

$p > .05, p = .14$ ). The results further revealed significant interaction effects between text structure and language proficiency indicating that different proficiency groups showed different performance patterns across different text genres.

Descriptive statistics for the performance of the three proficiency groups in the two tests are displayed in Table 4. In the FT recall test, there was no significant difference among the three proficiency groups ( $F(3, 2) = 2.66, p > .05$ ). In contrast, in the NT recall test, ANOVA revealed significant difference among the three groups,  $F(3, 2) = 15.14, p < .001$ . Scheffe multiple groups comparisons showed that the HP group outperformed the MP and the LP groups, whereas the performance of the latter two groups didn't differ significantly.

**Table 3. Two-way ANOVA summary for text structure and proficiency in the free-recall test**

Source	DF	SS	MS	F	P
Text structure	1	0.81	0.81	2.14	.144
Proficiency	2	12.77	6.39	16.97	.000*
Text structure x proficiency	2	3.90	1.95	5.19	.006*

Across different text structure, the HP group was found to score significantly higher in the NT,  $t = -3.43, p = .001$ . In contrast, the MP group performed equally poorly in the two recall tests,  $t = 0.30, p = .077$ ; whereas the LP group was found to perform significantly better in the FT,  $t = 2.08, p = .044$ .

**Table 4. Means and standard deviations of the scores for the free-recall test for HP, MP, and LP learner groups in the Fiction Task and Nonfiction Task**

Language group	n	FT structure		NT structure	
		M	SD	M	SD
High proficiency	40	1.35	0.66	1.82	1.15
Mid proficiency	39	1.17	0.45	1.15	0.43
Low proficiency	40	1.10	0.30	1.00	0.00
Total	119	1.21	0.50	1.32	0.79

Maximal score = 5

In summary, text structures and test formats were found to be factors affecting the subjects' performance. In a recognition task, subjects of different language proficiency levels performed significantly better in the FT than in the NT in the multiple-choice test ( $t = 6.04, p < .001$ ). In contrast, the HP performed significantly better in the NT in the recall test, the MP group performed equally poorly in the two recall tests, while the LP group was found to perform significantly better in the FT. This complies with the finding of Huang (2006) that written recall is a production task that demands a sustained cognitive effort in working memory.

## Conclusion

Text analyses showed that the selected fiction and nonfiction texts differed in word count, type-token ratio if text length was taken into account, and vocabulary difficulty.

Though the publisher judged the two texts to be of similar difficulty level by using various readability estimate formulas, empirical evidence provided by the present study, nonetheless, showed that the Chinese EFL 5th graders performed significantly better in comprehending the fiction text than the nonfiction text. Complied with the findings of previous studies conducted in L1 environment (Cheuvront, 2002; Dymock, 2005), the results indicated that for the Chinese EFL learners the nonfiction text also posed a great reading difficulty. Comparative studies examining the difficulty level of the pair books of the twin texts should be conducted in the future. In addition, reading instruction focusing on text structure should be implemented with effective teaching strategies, and grade-appropriate basal readers pairing fiction books with nonfiction books should be made available.

**Feng-Lan Kuo** received her Ph.D. degree in the Linguistics Department at University of Illinois at Urbana-Champaign in 1994. Her research interests include phonetics, phonology, phonics, and web-based English teaching and learning. She is currently serving as a professor in the English Department at National Changhua University of Education, Taiwan. <laflkuo@cc.ncue.edu.tw>

**Heien-Kun Chiang** completed his Ph.D. degree in the Computer Science Department at University of Illinois at Urbana-Champaign in 1997. His research interests include internet technology, e-learning, and intelligent tutoring and diagnosis systems. He is currently an associate professor in Information Management Department at National Changhua University of Education, Taiwan. <hkchiang@cc.ncue.edu.tw>

## Acknowledgement

This study is supported in part by National Science Council, Taiwan, under grant NSC 96-2411-H-018-009.

## References

- Alvermann, D. E., & Boothby, P. R. (1982). Text differences: Children's perceptions at the transition stage in reading. *The Reading Teacher*, 36(3), 298-302.
- Camp, D. (2000). It takes two: Teaching with twin texts of fact and fiction. *The Reading Teacher*, 53(5), 400-408.
- Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, 19(4), 727-752.
- Chevront, M. L. (2002). *Analysis of sensitivity and comprehension difficulty among expository text structures*. Unpublished doctoral dissertation, Boston University, Boston.
- Dymock, S. (2005). Teaching expository text structure awareness. *The Reading Teacher*, 59(2), 177-182.
- Flesch, R. F. (1949). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Halliday, M. A. K. (1985). *Spoken and written language*. Melbourne: Deakin University Press.
- Hess, C. W., Haug, H. T., & Landry, R. G. (1989). The reliability of type-token ratios for the oral language of school age children. *Journal of Speech and Hearing Research*, 32(3), 536-540.
- Huang, H. L. (2006). The effect of storytelling on EFL young learners' reading comprehension and word recall. *English Teaching and Learning*, 30(3), 51-74.
- Irwin, P. I., & Mitchell, J. N. (1983). A procedure for assessing the richness of retelling. *Journal of Reading*, 26(5), 391-396.
- Jeng, H. S. (2001). The CEEC English word list I and II. *The Proceedings of the Fifth International Conference on Multimedia Language Education*, 208-216. Taipei: Crane.
- Law, Y. K. (1995). Effects of text structures on recall. *Education Journal*, 23(1), 115-133. Retrieved March 26, 2008, from <http://sunzi1.lib.hku.hk/hkjo/article.jsp?book=33&issue=330032>
- Moss, B. (2004). Teaching expository text structures through information trade book retellings. *The Reading Teacher*, 57(8), 710-718.
- Schneider, P., & Vis Dub'e, R. V. (2005). Story presentation effects on children's retell content. *American Journal of Speech – Language Pathology Rockville*, 14(1), 52-60.
- Ure, J. N. (1971). Lexical density and register differentiation. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of Linguistics: Selected Papers of the 2nd International Congress of Applications of Linguistics* (pp. 443-452). Cambridge: Cambridge University Press.
- Ying, C. H. (2001). *Redundant elements in prepared speeches and simultaneous interpretation*. Unpublished master's thesis, National Taiwan Normal University, Taipei.
- Twin Texts (2002). [Online] Available: <http://www.sundancepub.com/c/@OROqxRteNHPuU/Pages/product.html?record@S3701>