

Scoring recalls for L2 readers of English in China: Pausal or idea units

Cindy Brantmeier & Michael Strube
Washington University in St. Louis
United States

Xiucheng Yu
Northeast Normal University
China

Abstract

Written recall may be a powerful tool used to address reading deficiencies in China. With 180 students enrolled in a third-year English class at a large university in northeastern China, the present investigation studies the relationship between pausal and idea units used to codify written recalls, and it investigates whether the strength of the relationship between pausal and idea units depends on other variables, such as length of time spent studying English, the amount of leisure reading done in English, or the version of passage. Findings indicate a strong correlation between idea units and pausal units for written recalls. This correlation underscores prior findings by Bernhardt (1991), and it reveals that the strength of the relationship between pausal and idea units does not depend on the moderating variables examined. Results are discussed in light of prior research and a detailed discussion of future directions for experiments of this type is offered.

Keywords: recall, units, comprehension, English

The demand for mastery of English in China has accelerated due to the demands to establish cultural, political, and economic relationships worldwide. More specifically, English is viewed as a powerful instrument necessary for the advancement of science and technology (Y. Gao, personal communication, February 20, 2013), and consequently China has the largest population of learners studying English in the world. Pang (2008) asserted that reading, as opposed to listening, speaking, or writing, is the skill most needed for both academic and professional reasons. While the ability to communicate orally in English is undeniably important to China, reading in English is also of vital significance so that China can continue its efforts to join the world economy. Traditionally, national reading examinations that are administered frequently by the Ministry of Education in China use multiple-choice tasks to test comprehension (Pang, 2008; X. Yu, personal communication, July 10, 2012; Zhang, 2010; Zhang & Wu, 2009) because it is more efficient to use computer-based scoring for large groups. Given the recent emphasis on second language (L2) reading research in China, Pang (2008) cautioned researchers about the

reliance on multiple-choice tests in China because it may impede the development and growth of the reader. With multiple-choice items, the questions and answers are predetermined, and consequently there is no room for the reader to generate inferences or think critically about the reading. Pang also included very specific directions for future research that include the examination of word-level issues, extensive reading, strategies, and testing. All of these suggestions for future research involve a measure of comprehension. Thus, the present study of learners of English in China examines the grading of an assessment task that is regarded by reading researchers in the USA (Bernhardt, 1991; Brantmeier, 2006) to be the most revealing of readers' comprehension: the written recall protocol.

The written recall protocol is a common procedure utilized to measure reading comprehension in both first language (L1) (Fitzgerald & Spiegel, 1983; Pearson & Camperell, 1981; Rand, 1984; Snyder & Downey, 1983) and second language (L2) investigations (Bernhardt, 1983; Brantmeier, 2002; Carrell, 1983a; Lee, 1986; Young & Oxford, 1997). In the written recall task students are asked to read a text and write down everything they can remember about what they just read without looking back at the text. Bernhardt (1991) initially postulated that the written recall was the purest measure of L2 reading comprehension, and to date this protocol continues to be used in many investigations that involve L2 reading comprehension. With the written recall there is no interference by the instructor who is proctoring the test and there are no retrieval cues provided.

Prior L2 Research and the Written Recall Protocol

Prior L2 research that examines variables involved in comprehension assessment tasks have primarily highlighted language of assessment as a key factor (Hock & Poh, 1979; Shohamy, 1984), and investigations have specifically examined language as a key variable in written recall protocol research (Brantmeier, 2006; Lee, 1986). Findings of these studies have guided important methodological decisions for empirical research on L2 reading. With students from beginning and intermediate levels of Spanish in the USA, Lee (1986) reported that written recall of a text was significantly better when completed in L1 rather than L2. Alderson (2000) also claimed that "recall should be completed in the test taker's L1 because otherwise it becomes a test of writing rather than reading" (p. 230). Bernhardt (2005) reinforced Alderson's recommendation, but only until learners reached the advanced L2 proficiency/fluency levels. With advanced learners of Spanish in the USA, Brantmeier (2006) investigated the extent to which L2 reading comprehension is a function of language of the recall task and/or former L2 reading performance. Findings indicated that overall language of recall does not matter with learners from advanced levels of language instruction. In that experiment, language of recall accounted for only 3% of variance in written recall. These findings underscore Bernhardt's (1991) earlier assertion that written recalls be completed in the native language at the beginning and intermediate stages of acquisition, but not at the advanced levels.

Review of codification schemes for written recall

In the two studies highlighted above that utilize learners of Spanish in the USA (Brantmeier, 2006; Lee, 1986), the scoring scheme used to codify written recalls was the pausal unit. Research with English as a Foreign Language (EFL) learners who are native Chinese speakers has also

explored factors that affect reading recall. One such study, by Chu, Swaffar and Charney (2002) examined the impact of culture-specific conventions on written recalls. They reported that different rhetorical conventions had significant effects on both immediate and delayed recalls. More specifically, researchers reported that most Taiwanese EFL students were not aware of cohesive devices when reading English texts as they occur less often in Chinese. English language learners in Taiwan rarely use cohesive devices for integrating textual information, and the authors contend that the difficulty in identifying cohesive ties and finding out the relationships among these devices in a text may negatively impact reading in English. In this study, the researchers also used the pausal unit to codify recalls. In all three studies (Brantmeier, 2006; Chu et al., 2002; Lee, 1986), researchers described the pausal unit as the natural pause during oral reading by a native speaker. In other words, a native speaker reads the passage out loud and inserts a bracket wherever there is a natural pause in speech. All information within a bracket is a pausal unit. This scoring process was first developed by Johnson (1970) and then validated by Bernhardt (1991) in a comparison of pausal units and the Meyer (1975) system, a protocol that includes hierarchically-based content structures. Bernhardt (1991) argued that both systems tap the same L2 reading behaviors, but that the pausal unit is more efficient and less time-consuming to grade. Consequently, many L2 reading researchers utilize the pausal unit as a system to score written recalls.

Lee and Ballman (1987) blended Johnson's (1970) definition of a pausal unit with Carrell's (1983b) notion of an idea unit, a unit that corresponds to a proposition or phrase, to divide an expository passage into units. They then ranked the importance of each unit. Their experiments specifically examined the effects of exposure to grammar on written recall. Findings revealed a low representation of important units in the L2 recalls. Since then, some L2 reading researchers have used a criterion from Riley and Lee (1996) whereby the text is divided into 'units of analysis,' which consist of ideas, propositions, and constituent structures. When dividing a reading into idea units, a small to medium grain size is used to determine the number of idea units within each passage. Each piece of information that adds meaning to the recall is classified as an idea unit. A list of idea units is agreed upon among raters. With this procedure, the entire written recall is analyzed by a researcher and an additional rater to identify the units, and then these units are compared to those of the test to ensure that the information in the written recall either appeared in the reading passage, or was implied in it. The total number of ideas recalled correctly is used for the recall score. Alderson (2000) pointed out that an idea unit was somewhat difficult to define, and that this issue has rarely been addressed in reading research. He was also careful to say that the objectivity of scoring with the idea unit depends on the rater reliability index and that high reliability is necessary.

One mechanism that merits further research is the scoring system used to codify written recalls, specifically with learners of English in China. To date, it appears that no study has examined the type of unit (pausal or idea) that is utilized to codify written recalls for native Chinese readers of English. The type of unit can also be described as the nature of the divisions within the written recall. The present study examines the relationship between the condition of unit (idea or pausal) under which recalls are scored. It also investigates the possible effects of moderating variables, such as the length of time spent studying English, the amount of pleasure reading done in English, and the type of passage used. These moderating variables were selected after discussions with the professors who teach the English courses at the university in China. These

professors spent one full year studying and doing research at a university in the USA, where they attended a course that examined the research and theory that drives L2 reading. After completing course readings on the interacting variables involved in reading, the professors felt that students who have studied English for a long time and frequently read in English for pleasure would not be affected by the condition of the unit. However, they also thought that the type of passage (passages with or without inserted questions) would impact recalls for all learners in the study. For the present study, the pausal unit is considered to be the natural pause produced by a native speaker while reading the passage orally (Bernhardt, 1991). The idea unit is an idea, proposition, or constituent structure (Riley & Lee, 1996).

In studies that use the pausal or idea unit protocol, the total number of correct units recalled is traditionally used as the dependent variable for comprehension. For the present investigation, the idea units are not weighted. In other words, no hierarchy was placed on the units recalled. The recalls were graded by counting the number of correct pausal units and the number of correct idea units.

The present study is motivated in part by prior findings from Brantmeier, Callender, Yu, and McDaniel (2012). In that study the written recalls were codified with the pausal unit protocol and were completed by two different researchers in China, both native speakers of Chinese and professors of English. The Chinese researchers had not used pausal unit in prior experiments, and after completing the scoring of all recalls they questioned the rubric because they thought it might have impacted scores. That investigation specifically examined the effects of textual enhancements on reading comprehension with adult native Chinese speakers learning English at the university. An adjunct was inserted into the readings that consisted of a ‘what’ question along with instructions to either ‘pause and consider,’ or ‘pause and write.’ Participants achieved almost the same scores on recall and sentence completion for versions with no adjuncts and versions with ‘pause and consider’ adjuncts, and they scored significantly lower on the assessment tasks for the passage versions with pause and write adjuncts. Could the scoring system used to codify recalls have made a difference in the outcomes? In other words, if recalls from that study had been codified with idea units, would different findings have emerged? The present study used a subset of data from the prior experiment to answer the following research questions. The prior investigation did not specifically address the research questions in the present study, as these new questions emerged after making methodological changes to the earlier study.

Research Questions

The following overall research questions guide the present study:

1. What is the relationship between pausal and idea units used to codify the written recalls of high-intermediate learners of English whose native language is Chinese?
2. Does the strength of the relationship between pausal and idea units depend on other variables, such as the length of time studying in English, the amount of pleasure reading done in English, or the version of passage used?

Method

Participants

Participants included 185 students (aged 19 to 24 years) enrolled in a third-year English course at a large university in the northeastern region of China. All participants (24 male and 161 female) declared English as their primary major. All students enrolled in this course participated in the experiment, and therefore the gender of the participants is not proportional. The English language program classified the learners as upper-intermediate, and all participants were enrolled in a required Advanced Reading course for English majors. The course goals involved development of the ability to analyze and paraphrase a variety of readings that included passages from professional textbooks as well as authentic texts written by native English speakers for English speakers. Through the introduction of different cultural knowledge, the course also examined rhetorical devices and writing features. Only those students who met the following criteria were included in the final data analysis: (a) students whose native language was Chinese, and (b) students who completed all tasks during data collection.

During their previous two years at the university, participants had been enrolled in various reading-related classes, such as Intensive Reading, Extensive Reading, and Journal Reading. In all these courses, students were required to read a lot outside of class. A reading list was provided before each semester, and multiple-choice tests of vocabulary and comprehension were conducted once a month to test for comprehension. Students were not familiar with the written recall protocol to measure reading comprehension. It is important to note that before enrolling in the course, all participants were required to have passed the national proficiency test called TEM-4 (Test for English Majors, Band 4) administered by the Ministry of Education of China.

Procedures

Data collection procedures. Data collection took place toward the end of the fall semester of 2010. The university is known for the outstanding quality of its language programs. Participants attending the Advanced Reading course were invited to take part in the experiment. Everyone agreed and then signed a consent form to that effect. Students had the option of extra credit points or monetary compensation for participation, and all participants preferred points.

When participants were invited to participate in the experiment, the researcher told them that the investigation would examine variables involved in the reading of English, but no specific details were provided. The researcher was present during data collection to ensure that students did not turn back to previous pages while completing the tasks. The experiment was held on a single day outside of class time.

Demographic Questionnaire. A demographic questionnaire was developed in order to examine other factors that may be involved in L2 reading comprehension as measured via the written recall protocol. Questions such as ‘What is your native language?’ and ‘How many years have you studied English?’ were offered as part of the questionnaire. The complete questionnaire is included in Appendix A. As indicated earlier, participants completed the demographic questionnaire prior to the readings as part of the data collection packet. Data was then entered

directly into the SPSS database as the questions were direct and no codification process was necessary.

Reading Passages and Embedded Questions. The two reading passages for the study were taken from social psychology textbooks and from passages utilized in prior L1 reading research (Callender & McDaniel, 2007) and L2 reading research (Brantmeier et al., 2012). The first passage included details about theories of first impressions, the primacy effect, and schemas. The second passage detailed implicit personality theory and attribution theory. Table 1 lists the passage length, total number of sentences, and total number of embedded clauses, and also includes factors related to text difficulty. For the sake of consistency, the authors of the present study included the above factors to determine passage difficulty (Brantmeier et al., 2012). A future study of this type may also analyze text difficulty with Flesch Reading Ease scales to further analyze the effects of text difficulty.

Table 1. *Text difficulty by passage*

Passage Content	Length	Number of Sentences	Number of Embedded Clauses
Passage 1: First impressions, the primacy effect, and schemas	525 words	21 sentences	15 embedded clauses
Passage 2: Implicit personality theories with a detailed explanation of attribution theory	646 words	22 sentences	16 embedded clauses

Appendix B includes the breakdown of pausal and idea units for each passage, and it includes both the English and Chinese translations. Passage 1 the total number of pausal units was 26 and total number of idea units was 52. For Passage 2 the total number of pausal units was 24 and the total number of idea units was 48. It is important to note that generally speaking, the breaks for pausal units do not occur in the same locations as the breaks for idea units. Three different versions of each passage were included in the study (Brantmeier et al., 2012). Version 1 included no inserted questions. Version 2 included inserted ‘what’ questions that required learners to pause and consider, and Version 3 included inserted ‘what’ questions that asked learners to pause and write.

Recall. As discussed in detail earlier, the instructions for the written recall asked: ‘Please write down as much as you can remember about what you just read.’ The third author of this study, a native speaker of Mandarin Chinese, scored recalls first for the presence or absence of pausal units, then for the presence or absence of idea units. For inter-rater reliability, a second native Chinese speaker was trained to score recalls for the two types of units, and then randomly scored written recalls. The inter-rater reliability coefficient was .97 for pausal units and .96 for idea units. The total number of possible pausal units was 26 for Passage 1, and 24 for Passage 2. The total number of possible idea units for was 51 for Passage 1, and 48 for Passage 2.

Results

Means, standard deviations, and correlations for total pausal and idea units recalled for the combined readings are listed in Table 2. Across passages, participants scored higher on the

recalls that were codified according to idea units than they did on the recalls that used pausal units. Recall across passages was strongly correlated for both pausal units and idea units; within passages, pausal and idea unit recall was even more strongly correlated. To simplify subsequent analyses, the pausal unit recall scores for the two passages were averaged (after standardization). Similarly, the idea unit recall scores for the two passages were also averaged. The percentage of correct pausal units was 81% for Passage 1 and for 82% for Passage 2. The percentage of correct idea units was 75% for for Passage 1 and 65% for Passage 2.

Table 2. Means, standard deviations and correlations for pausal and idea unit recall

		Pausal Units		Idea Units	
		Passage 1	Passage 2	Passage 1	Passage 2
Pausal Units	Passage 1	1.00			
	Passage 2	0.66	1.00		
Idea Units	Passage 1	0.94	0.68	1.00	
	Passage 2	0.66	0.95	0.70	1.00
<i>Mean</i>		21.37	19.97	37.92	30.81
<i>SD</i>		11.90	11.97	19.34	17.31

Note. All correlations are statistically significant, $p < .001$.

To illustrate the strong linear relationship between pausal unit and idea unit recall, Figure 1 shows the scatter diagram for the averaged data. As depicted on Figure 1, the relationship is very high, with over 90% of the variance in pausal unit recall accounted for by idea unit recall.

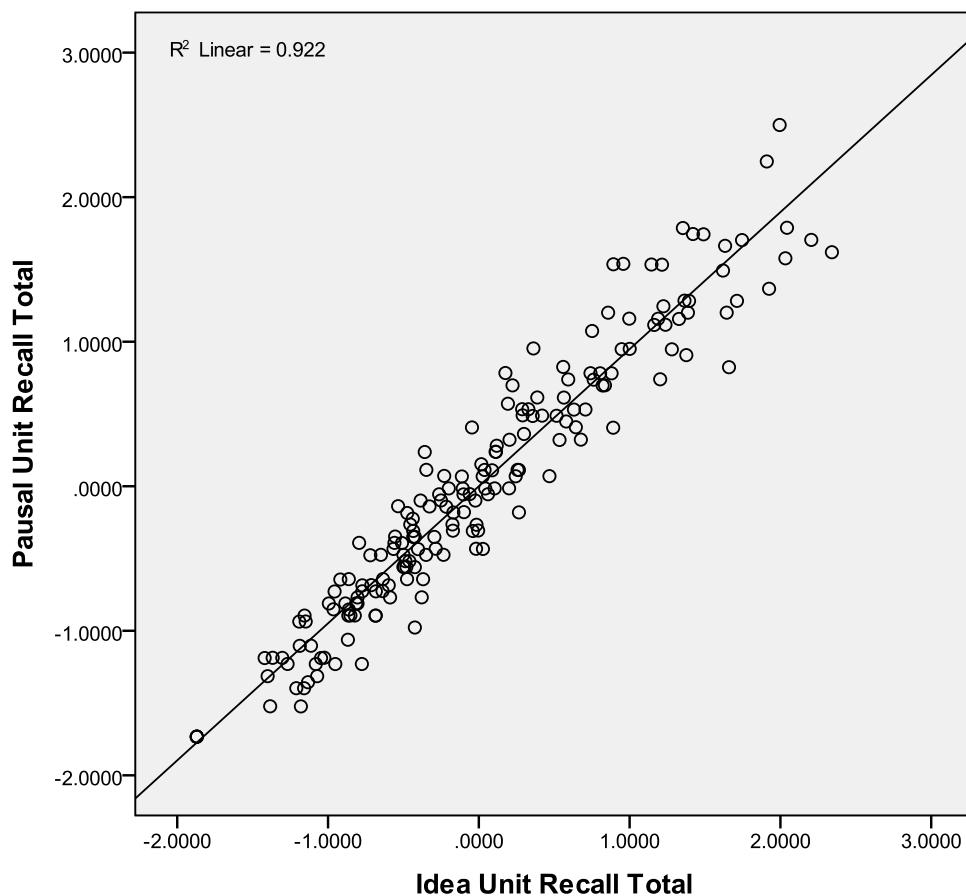


Figure 1. Relationship between pausal and idea units

A close inspection of the scatterplot reveals somewhat greater deviations from the regression line near the upper range of recall, suggesting the possible presence of moderating variables. To explore this possibility, multiple regression analyses were conducted with total pausal unit recall as the outcome, total idea unit recall as one predictor, an additional variable as another predictor (e.g., length of time studying in English, amount of leisure reading done in English, or version of passage read), and the product of idea unit recall and the additional predictor included to test the interaction (i.e., presence of a moderator). The only significant factor that emerged was version of passage read, $F(2, 179) = 4.46, p = .013$. The factors of length of study and amount of pleasure reading did not emerge as significant variables in the present study. Figure 2 illustrates the slightly different relations between pausal unit recall and idea unit recall for the three different versions, with slight attenuation for Version 1 (no inserted questions) and Version 2 (pause and consider) relative to Version 3 (pause and write). Note, however, that these differences are rather small amid a large relationship between the two recall measures ($r = .96$).

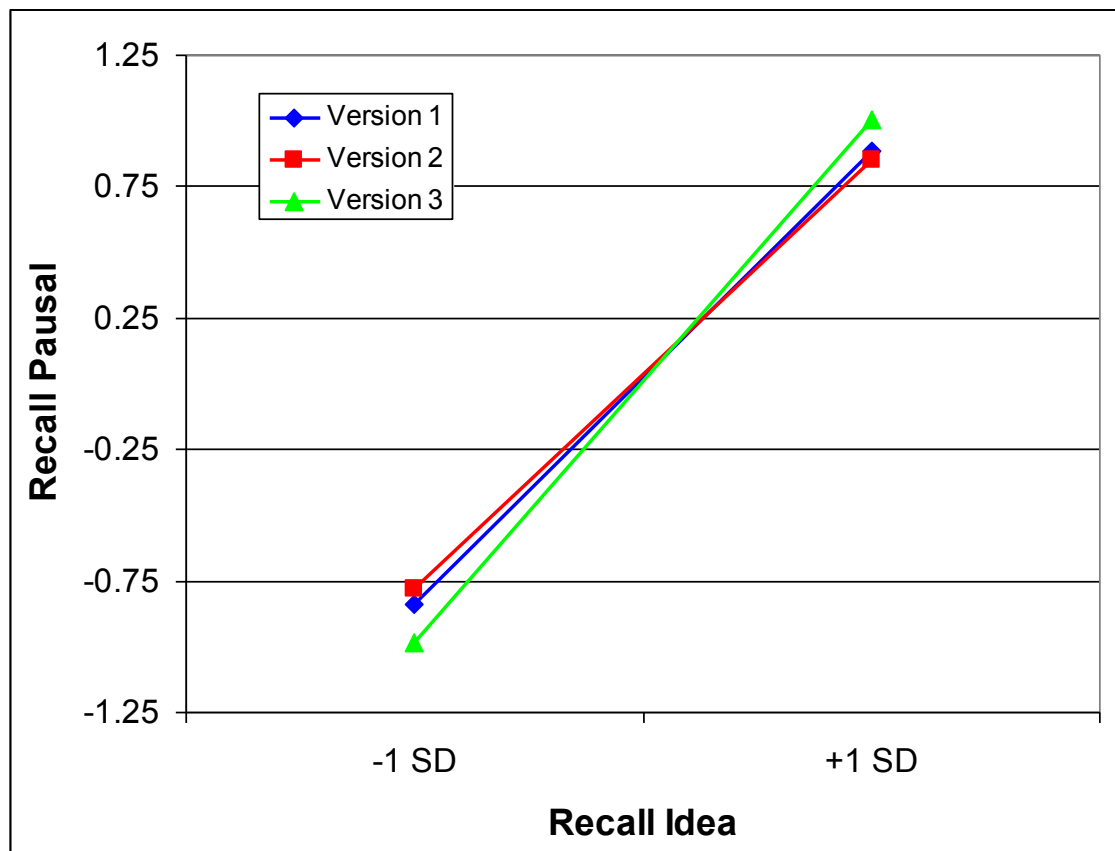


Figure 2. The effect of version on the relation between idea unit recall and pausal unit recall

Discussion

Overall, the present study found a strong correlation between idea units and pausal units for written recalls. This correlation appears to underscore prior findings by Bernhardt (1991) with native English speakers learning to read in German. Bernhardt provides examples of a weighted propositional analysis of recalls, the Meyer system, and compares this to the simple pausal unit analysis of recalls. In her study, correlations between the Meyer system and the pausal system were .96 for one text. In the second text, however, the correlation was only .54. With the second text Bernhardt followed with additional weighted scoring that brought the correlation up to .84. In the end, Bernhardt demonstrated that the pausal unit was just as effective as the idea unit, and that it was far more efficient to use it, as it was less time consuming for the researchers. This was also the case for the present experiment with native Chinese learners of English. The strong correlation between idea and pausal unit scores in the present study underscores Bernhardt's (1991) earlier assertion that the pausal and idea units tap the same behavior.

More specifically, even though the number of possible idea units for each passage was higher than the number of possible pausal units, the findings echoed Bernhardt (1991). The present study did not scale or rate the types of idea units recalled, and it did not find that the idea units were a better system for grading. When they were completing coding the recall tasks, the Chinese researchers questioned the use of pausal units and asked if idea units would better serve

these readers. The researchers were not accustomed to using the written recall protocol as a measure of comprehension, probably because multiple-choice questions are more commonly used in China. The findings demonstrate that the pausal unit grading system is appropriate for English learners in China.

It is important to discuss the fact that although the high correlation between the two measures suggests that they are interchangeable, this is only true in a relative sense, not in an absolute sense. The percentage of correct idea units was 75% for Passage 1 and 65% for Passage 2. The idea unit approach seems to produce systematically lower scores. A follow-up study could examine more qualitative information within recalls, such as misunderstandings, and how these misunderstandings impact comprehension. For example, Young and Nakuma (2009) used a subset of data from an earlier experiment (Young, 1999) to examine the misunderstandings present in recalls of Spanish learners. Their second-year Spanish students read culturally unfamiliar texts and were asked to recall as much as they could about what they had read in their native language. The researchers categorized the incorrectly recalled units as either linguistically-based misunderstandings, where students' inadequate knowledge of the language code may explain misunderstandings, or cognitively related misunderstandings that are attributable to propositional deficits. The misunderstandings consisted of students rewriting the text to fit their personal assumptions. Findings revealed that when readers at this stage of acquisition read a text with little background knowledge, they became more dependent on the literal language of the text. The researchers found that the readers used more word-level approaches to comprehend and that their linguistic deficits also strongly influenced misunderstandings of the reading passages. The recalls provided in the present study could be examined for misunderstandings with a linguistically-based and cognitively-related rubric in order to find out whether the students of English in China were using this type of compensatory behavior because of deficiencies in linguistic abilities or lack of familiar content located within domain-specific texts. The result of such analysis could hold strong implications for the teaching of reading in China because the nature of misunderstandings may be more easily revealed with one type of unit than the other. The present study indicated that from an overall performance standpoint it does not matter which type of grading unit is used, but it may matter for a more fine-grained analysis. This awaits additional research to resolve.

Brantmeier et al. (2012) discussed reasons for decreased performance in the 'pause and write' condition by explaining that the question and writing process may have caused the readers to focus on the information addressed in the question, rather than generating inferences and other details from the text. The questions for 'pause and write' were predetermined retrieval cues. Consequently, the readers may have thought this information was of most importance. As inferences are included in the idea unit rubric but not on the pausal unit protocol, we could assert that the embedded adjuncts did not help the reader with the idea units. When writing the recall the readers may have been influenced by the nature of these questions. A close examination of the recalls revealed that participants did not include other information from the text; this may have had an impact on the nature of units recalled.

Scholars and instructors should be cautious about how far the current findings are to be applied. The results indicated that pausal units are indeed a fine proxy for idea unit recall and are easier to use. However this finding applied to the overall performance on the recall task. Future research

should determine if the pausal and idea unit grading rubrics are differentially sensitive to particular kinds of recall errors.

Finally, the strength of the relationship between pausal and idea units does not depend on the moderating variables examined in this study. The current investigation reveals a moderating effect for version of passage, but it was very small relative to the high correlation between idea unit recall and pausal unit recall. In the context of the regression analysis, one reason for the significant moderating effect of version of passage is that the error term used to test that effect is reduced by removing the sizeable main effect of idea unit recall. More specifically, for versions 1, 2, and 3 the correlations between total pausal unit recall and total idea unit recall were .956, .958, and .943 respectively. Although the moderating effect of version of passage indicates these relationships are statistically different (technically that the regression coefficients are different), those differences are clearly quite small relative to the very substantial magnitude of the correlations. It is important to note that the main finding of the present study is that, when looking at overall performance, the pausal and idea units of analysis both tap into the same reading ability.

Conclusion

The findings of the present study provide support for the use of pausal units to grade written recalls. Further research should address how the pausal unit codifying scheme may help to identify reading deficiencies with consequential implications for instruction. The present study is the first to address the use of these two commonly-used tools, pausal and idea units, for grading in China. This study adopted a quantitative approach to data collection, but in the future it will be important to undertake qualitative investigations for more in-depth exploration of this phenomenon. Such studies could offer implications for reading material developers and test constructors in China. For example, written recalls could be used during class time where the students read the passage, recall what they just read, and then grade recalls with a pausal unit template. Students could follow-up with pair work where they compare written recalls and take a close look at the quality of the pausal units. Authors of textbooks for China may want to consider these activities as a part of their textbook design.

References

- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Bernhardt, E. B. (1983). Syntactic and lexical/semantic skill in foreign language reading comprehension: The immediate recall protocol. *Die Unterrichtspraxis*, 16, 27–33.
- Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, research and classroom perspectives*. Norwood, NJ: Albex.
- Bernhardt, E. B. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics*, 24, 133–150. doi: 10.1017.S0267190505000073
- Brantmeier, C. (2002). The effects of passage content on second language reading comprehension by gender across instruction levels. In J. Hammadou-Sullivan (Ed.), *Research in second language learning: Literacy and the second language learner* (pp.

- 149–176). Greenwich, CT: Information Age Publishing.
- Brantmeier, C. (2006). The effects of language of assessment and L2 reading performance on advanced readers' recall. *The Reading Matrix*, 6, 1–17.
- Brantmeier, C., Callender, A., Yu, X., & McDaniel, M. (2012). Textual enhancements and comprehension with adult readers of English from China. *Reading in a Foreign Language*, 24, 158–184.
- Callender, A. A., & McDaniel, M. A. (2007). The benefits of embedded questions for low and high structure builders. *Journal of Educational Psychology*, 99, 339–348.
doi:10.1037/0022-0663.99.2.339
- Carrell, P. L. (1983a). Some issues in studying the role of schemata, or background knowledge, in second language comprehension. *Reading in a Foreign Language*, 1, 81–92.
- Carrell, P. L. (1983b). Three components of background knowledge in reading comprehension. *Language Learning*, 33, 183–203. doi: 10.1111/j.1467-1770.1983.tb00534.x
- Chu, H. C. J., Swaffar, J., & Charney, D. H. (2002). Cultural representations of rhetorical conventions: The effects on reading recall. *TESOL Quarterly*, 36, 511–541.
doi: 10.2307/3588239
- Fitzgerald, J., & Spiegel, D. L. (1983). Enhancing children's reading comprehension through instruction in narrative structure. *Journal of Reading Behavior*, 15(2), 1–17.
doi: 10.1080/10862968309547480
- Hock, T. S., & Poh, L. C. (1979). The performance of a group of Malay-medium students in an English reading comprehension test. *RELC Journal*, 10, 81–89.
doi: 10.1177/003368827901000106
- Johnson, R. (1970). Recall of prose as a function of the structural importance of linguistic units. *Journal of Verbal Learning and Linguistic Behavior*, 9, 12–20. doi: 10.1016/S0022-5371(70)80003-2
- Lee, J. F. (1986). On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition*, 8, 201–212. doi: 10.1017/S0272263100006082
- Lee, J. F., & Ballman, T. L. (1987). Learners' ability to recall and rate important ideas of an expository text. In B. Van Patten, T. T. Dvorck, & J. F. Lee (Eds.), *Foreign language learning: A research perspective* (pp. 108–117). Rowley, MA: Newbury House.
- Meyer, B. J. F. (1975). *The organization of prose and its effect on memory*. Amsterdam: North Holland Publishing Company.
- Pang, J. (2008). Research on good and poor reader characteristics: Implications for L2 reading research in China. *Reading in a Foreign Language*, 20, 1–31.
- Pearson, P. D., & Camperell, K. (1981). Comprehension of test structures. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 27–55). Newark, DE: International Reading Association.
- Rand, M. K. (1984). Story schema: Theory, research and practice. *The Reading Teacher*, 37, 377–382.
- Riley, G. L., & Lee, J. F. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. *Language Testing*, 1, 173–187.
doi: 10.1177/026553229601300203
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Reading Research Quarterly*, 17, 229–255.
doi: 10.1177/026553228400100203
- Snyder, L., & Downey, D. (1983). Pragmatics and information processing. *Topics in Language*

- Disorders*, 4, 75–86.
- Young, D. J. (1999). Linguistic simplification of SL reading material: Effective instructional practice? *The Modern Language Journal*, 83, 350–366. doi: 10.1111/0026-7902.00027
- Young, D. J., & Nakuma, C. (2009). What don't you understand? Understanding misunderstandings in foreign language reading. In C. Brantmeier (Ed.), *Crossing languages and research methods: Analyses of adult foreign language reading*, (pp. 73–96). Charlotte, NC: Information Age Publishing.
- Young, D. J., & Oxford, R. (1997). A gender-related analysis of strategies used to process input in the native language and in a foreign language. *Applied Language Learning*, 8, 43–73.
- Zhang, L. J. (2010). A dynamic metacognitive systems account of Chinese university students' knowledge about EFL reading. *TESOL Quarterly*, 44, 320–353. doi: 10.5054/tq.2010.223352
- Zhang, L. J., & Wu, A. (2009). Chinese senior high school EFL students' metacognitive awareness and use of reading strategies. *Reading in a Foreign Language*, 21, 37–59.

Appendix A

Demographic Questionnaire

1. Age: _____
2. Male: _____ Female: _____
3. What is your native language? _____
4. How many years have you studied English? _____
5. Where did you study English? ___ elementary school ___ secondary school ___ private language center
6. How long have you studied English at your university? _____
7. Have you ever been to an English speaking country? _____ For how long? _____ Where? _____
8. How will you use English in your future? _____
9. What do you plan to do when you graduate from the university? _____
10. What skill (listening, speaking, reading, or writing) do you think is the most important to learn in English? ___ Why?
11. Do you read in English for pleasure? ___ If so, what do you read?
12. How often do you read in English? _____
13. How would you rate your ability to read in English? _____
 - a. Excellent
 - b. Very good
 - c. Ok
 - d. Not very good
 - e. Not very good at all

Appendix B

Pausal and Idea Unit Breakdown by Passage

Pausal Units - Each unit is indicated with a bracket /

FIRST IMPRESSIONS

First impressions are the initial judgments we make about people,¹ and they play an important role in social perceptions.² We are more likely to form opinions of others quickly³ based on first impressions⁴ than to refrain from forming opinions until we have more information.⁵ These first impressions may change as we get to know a person better⁶ but we often tend to hang on to them even in the face of contradictory evidence.⁷ Thus, initial opinions may have a strong impact on our future interactions with people.⁸

第一印象是我们对他人的最初评价，¹ 它们在社会认知中起到很重要的作用² 基于第一印象³ 我们会比等到有了更多信息才形成对他人的观点⁴ 更可能很快形成对他人的观点⁵ 这些第一印象会随着我们对一个人的了解的加深而改变⁶ 但我们即使在有了相反的证据面前通常还坚持我们的第一印象⁷ 因此，最初的观点可能会对我们在将来与人交往有重要的影响⁸ (8)

For example,¹ if you first meet a new tenant in your apartment building at a party at which he appears to be loud and egotistical² it will probably be hard for you to perceive him as a sensitive, caring person³ when you later see him comforting a small child who has scraped his knee⁴ The first information we receive about a person often seems to count the most⁵ a phenomenon referred to as the **primacy effect**.⁶ (6)

譬如¹ 你在一个晚会上第一次遇到了和你同住一栋楼的住户，在晚会上他表现得很吵闹，并以自我为中心² 这样当后来你看到他在安慰一个擦破了膝盖的小孩时³ 你就很难把他视为一个敏感、体贴的人⁴ 关于一个人最初的信息通常显得最重要⁵ 这种现象被称为首因效应⁶

This effect was demonstrated in an experiment¹ in which two lists of traits describing a person were read to two separate groups of subjects² (Asch, 1946). In one group³ subjects heard a description that began with positive characteristics⁴ (such as intelligent and industrious)⁵ followed by negative ones⁶ (impulsive, stubborn, and so forth).⁷ Their overall assessments of this person were positive.⁸ Subjects in the other group heard the same list⁹ but in reverse order.¹⁰ The results:¹¹ Their assessments were far more negative.¹² (12)

有一个实验展示了这种效应¹ 在这个实验中，两组用来描述一个人的特性被分别读给两组受试² 在一个组中³ 受试先听到的是关于人的积极特点的描述⁴ (如聪明、勤奋) ⁵ 接下来才听到消极特点的描述⁶ (冲动、固执，等等) ⁷ 受试对这个人的总体评价是积极的⁸ 另外一组的受试听到的描述特点是一样的⁹ 但是顺序相反¹⁰ 结果是：¹¹ 他们对这个人的评价就要负面的多¹²

26 total possible pausal units

SCHEMAS

Schemas are the conceptual frameworks we use to make sense out of our world.¹ The concept of schemas helps explain how we perceive the people we meet.² For example³ you might have schemas of lawyers as aggressive and intelligent,⁴ and of professor as studious and somewhat introverted.⁵ Social psychologists refer to these generalized assumptions about certain classes of people as *person schemas*.⁶ (6p)

图式是我们理解外部世界的概念框架¹ 图式的这个概念帮助我们解释我们如何认知我们所遇到的人² 比如³ 你对律师的图式可能是咄咄逼人和聪明⁴ 对于教授的图式是勤奋并且有点内向⁵ 社会心理学家把这些关于某些类型的人的概括的假设称为个人图式⁶

Person schemas provide a structure for evaluating the people we meet¹, allowing us to take shortcuts by concentrating on some facts and ignoring others². When we assess others for the first time³ we tend to pick up only the information that fits our existing schemas⁴ ignoring the rest.⁵ This process is cognitively efficient,⁶ but, unfortunately,⁷ it is not always the most accurate way of forming impressions⁸ (Brigham, 1986).

个人图式为我们评价遇到的人提供了一个结构¹ 使我们能够走捷径关注一些事实而忽略另外一些事实² 当我们第一次评价他人时³ 我们会刻意选择那些符合我们的图式的信息⁴ 而忽略其余的信息

⁵ 这个过程在认知上是又效率的⁶ 但遗憾的是⁷ 它不总是形成印象的最准确的方式⁸ (8p)
 Once we fit a person into a schema¹, we tend to use that schema as a general organizing principle² for interpreting further information about the person³. For example⁴, if our first impression of a new neighbor is that she is unfriendly⁵, we are likely to evaluate her failure to comment on our shiny new car as further evidence for unfriendliness⁶. If she then acts in a way that does not fit the schema⁷ (for example, picking up our garbage after it has been scattered by a storm)⁸, we may dismiss that act⁹ by concluding that she picked up the mess only because she was worried that it would blow onto her lawn.¹⁰
 一旦我们把一个人框在一个图式中¹ 我们会把那个图式作为一个总的组织原则² 来解释对于那个人进一步的信息³ 例如⁴ 如果我们对于一个新的邻居的第一印象时她不怎么友好⁵ 我们很可能把她未能夸赞我们崭新的汽车作为不友好的进一步的证据⁶ 而如果她的行为和我们的图式不符⁷ (例如, 在一场暴风雨过后捡起四散的垃圾) ⁸ 我们会对那种行为不予考虑⁹ 而会解释说她收拾垃圾仅仅时因为她担心会被吹到她的草坪上去¹⁰ (10p)

24 total possible pausal units

Idea Units

FIRST IMPRESSIONS

First impressions¹ are the initial judgments² we make about people³, and they play an important role⁴ in social perceptions⁵. We are more likely to form opinions⁶ of others quickly⁷ based on first impressions⁸, than to refrain from forming opinions⁹ until we have more information¹⁰. These first impressions may change¹¹ as we get to know a person better¹², but we often tend to hang on to them¹³ even in the face of contradictory evidence¹⁴. Thus, initial opinions may have a strong impact¹⁵ on our future interactions with people.¹⁶ (16 idea units)

第一印象¹ 是我们对他人的² 最初评价, ³ 它们在社会认知中⁴ 起到很重要的作用⁵。基于第一印象⁶, 我们会比等到有了更多信息⁷ 才形成对他人的观点⁸ 可能很快⁹ 形成对他人的观点¹⁰。这些第一印象会随着我们对一个人的了解的加深¹¹ 而改变¹²。但我们即使在有了相反的证据面前¹³ 通常还坚持我们的第一印象¹⁴。因此, 最初的观点可能会对我们在将来与人交往¹⁵ 有重要的影响¹⁶ (16 idea units)

For example, if you first meet a new tenant¹ in your apartment building² at a party³ at which he appears to be loud⁴ and egotistical⁵, it will probably be hard for you⁶ to perceive him⁷ as a sensitive⁸, caring person when you later see him⁹ comforting a small child¹⁰ who has scraped his knee¹¹. The first information we receive about a person¹² often seems to count the most¹³, a phenomenon¹⁴ referred to as the **primacy effect**¹⁵. (15)

譬如, 你在一个晚会上¹ 第一次遇到了² 和你同住一栋楼的住户³, 在晚会上他表现得很吵闹⁴, 并以自我为中心⁵, 这样当后来你看到他⁶ 在安慰一个擦破了膝盖的小孩时⁷, 你就很难⁸ 把他视为⁹ 一个敏感¹⁰、体贴的人¹¹。关于一个人最初的信息¹² 通常显得最重要¹³ 这种现象¹⁴ 被称为首因效应¹⁵。(15)

This effect was demonstrated¹ in an experiment² in which two lists of traits³ describing a person⁴ were read to two separate groups of subjects⁵ (Asch, 1946). In one group⁶, subjects heard a description⁷ that began with positive characteristics⁸ (such as intelligent⁹ and industrious¹⁰) followed by negative ones¹¹ (impulsive¹², stubborn¹³, and so forth¹⁴). Their overall assessments¹⁵ of this person were positive¹⁶. Subjects in the other group¹⁷ heard the same list¹⁸, but in reverse order¹⁹. The results²⁰: Their assessments were far more negative²¹. (21)

有一个实验¹ 展示了这种效应²。在这个实验中, 两组用来描述一个人³ 的特性⁴ 被分别读给两组受试⁵。在一个组中⁶, 受试先听到的是⁷ 关于人的积极特点的描述⁸ (如聪明⁹、勤奋¹⁰), 接下来才听到消极特点的描述¹¹ (冲动¹²、固执¹³, 等等¹⁴), 受试对这个人的总体评价¹⁵ 是积极的¹⁶; 另外一组的受试¹⁷ 听到的描述特点是一样的¹⁸, 但是顺序相反¹⁹, 结果是²⁰: 他们对这个人

的评价就要负面的多²¹

52 total possible idea units 51 total possible idea units

SCHEMAS

Schemas are the conceptual frameworks¹ we use to make sense out of our world². The concept of schemas³ helps explain how we perceive the people⁴ we meet⁵. For example, you might have schemas of lawyers⁶ as aggressive⁷ and intelligent⁸, and of professors⁹ as studious¹⁰ and somewhat introverted.¹¹ Social psychologists¹² refer to these generalized assumptions¹³ about certain classes of people¹⁴ as *person schemas*.¹⁵

图式是我们理解外部世界¹的概念框架²。图式的这个概念³帮助我们解释我们如何认知⁴我们所遇到的人⁵比如你对律师的图式⁶可能是咄咄逼人⁷和聪明⁸对于教授的图式⁹是勤奋¹⁰并且有点内向¹¹社会心理学家¹²把这些关于某些类型的人¹³的概括的假设¹⁴称为个人图式¹⁵ (15)

Person schemas provide a structure¹ for evaluating the people² we meet,³ allowing us to take shortcuts⁴ by concentrating on some facts⁵ and ignoring others⁶. When we assess others⁷ for the first time⁸, we tend to pick up only the information that fits our existing schemas⁹, ignoring the rest¹⁰. This process¹¹ is cognitively efficient,¹² but, unfortunately, it is not always the most accurate¹³ way of forming impressions¹⁴ (Brigham, 1986).

个人图式为我们评价¹遇到的人²提供了一个结构³,使我们能够走捷径⁴关注一些事实⁵而忽略另外一些事实⁶当我们第一次⁷评价他人时⁸我们会刻意选择那些符合我们的图式的信息⁹而忽略其余的信息¹⁰这个过程¹¹在认知上是又效率的¹²但遗憾的是,它不总是形成印象的¹³最准确的方式¹⁴ (14)

Once we fit a person into a schema¹, we tend to use that schema as a general organizing² principle³ for interpreting further information⁴ about the person⁵. For example, if our first impression of a new⁶ neighbor⁷ is that she is unfriendly⁸, we are likely to evaluate her failure to comment⁹ on our shiny new car¹⁰ as further evidence for unfriendliness.¹¹ If she then acts¹² in a way that does not fit the schema¹³ (for example, picking up our garbage¹⁴ after it has been scattered by a storm¹⁵), we may dismiss that act¹⁶ by concluding that she picked up the mess /only because she was worried that it would blow onto her lawn.

一旦我们把一个人框在一个图式中¹我们会把那个图式作为一个总的组织²原则³来解释⁴对于那个人进一步的信息⁵例如,如果我们对于一个新的⁶邻居⁷的第一印象时她不怎么友好⁸我们很可能把她未能夸赞⁹我们崭新的汽车¹⁰作为不友好的进一步的证据¹¹而如果她的行为¹²和我们的图式不符¹³(例如,在一场暴风雨过后¹⁴捡起四散的垃圾¹⁵),我们会对那种行为不予考虑¹⁶而会解释说她收拾垃圾仅仅时因为她担心会被吹到她的草坪上去/

48 possible idea units

About the Authors

Cindy Brantmeier is associate professor of Applied Linguistics and Education in the Department of Education at Washington University in St. Louis. She is principal investigator in the Language Research Lab, co-director of the Graduate Certificate in Language Instruction, and director of Applied Linguistics. Dr. Brantmeier has published articles concerning interacting variables in adult second language reading, language research methodology, testing and assessment, and other related areas in journals such as *The Modern Language Journal*, *System*, *Foreign Language Annals*, and *Reading in a Foreign Language*, among others. She is also editor of several volumes and a book on adult foreign language reading. E-mail: cbrantme@wustl.edu

Michael Strube is professor of Psychology at Washington University in St. Louis. His research interests include self-knowledge, self-motives, person-environment fit, health psychology, research design, and statistics. He is a past editor of *Basic and Applied Social Psychology*. E-mail: mjstrube@wustl.edu

Xiucheng Yu was a research associate at Washington University in St. Louis from 2009 to 2010, where he collaborated with Professor Cindy Brantmeier in the Language Research Lab. He is now associate professor of English and English linguistics at Northeast Normal University (NENU) in China. Before teaching English majors at NENU, he taught ESP at Norman Bethune University of Medical Sciences and Changchun Taxation College. Dr. Yu offers courses such as Business English (Reading), Grammar and Writing, English Lexicology, and General Linguistics to undergraduates. He also teaches General Linguistics (Advanced) and Linguistics and Foreign Language Teaching to graduate students. His research interests mainly involve FLT, pragmatics (especially speech act studies), and cross-cultural communication. Professor Yu also serves as coordinator of the Writing Center at NENU. He is now deputy director of the English Department. E-mail: yuxc765@nenu.edu.cn