Title: Variations in the analysis of written recall protocols

Author: Yuh-Fang Chang

Publication date:4/8/2009

Abstract

While researchers generally quantify the amount of information that learners recall correctly in order to measure reading comprehension, the unit of analysis adopted to score the recall protocol differs.   Whether and how different scoring systems bring about a different picture of L2 reading comprehension, however, remains unexplored.   This study attempted to: 1) identify the commonly used scoring/analysis systems in the L2 reading literature and 2) make an in-depth comparison among the scoring systems.   The results revealed that the recall scores generated by the three systems are highly correlated even though the principles behind each scoring system differ.   These three scoring systems, however, differ in how straightforward and objective they are in allowing researchers to judge the correctness of learner's responses in their recalls.

Keywords: immediate written recall task, variations, Johnson's system, Bovair and Kieras's system, idea unit system

Introduction

The immediate written recall task has been claimed to be one of the most frequently used reading comprehension measures (Davis, 1989). The scoring of the recall protocols requires the researcher to first divide the reading text into a unit of analysis and then calculate the total number of units recalled correctly. Over the past decades, several scoring/analysis systems with different units of analysis have been developed. While researchers generally quantify the amount of information that learners recall correctly in order to measure reading comprehension, the unit of analysis adopted to score the recall protocol differs. Whether and how different scoring systems bring about a different picture of L2 reading comprehension, however, remains unexplored. To fully understand the variation in the analysis of the recall protocols is essential because it allows reliable comparisons to be made across research findings. The comparability of findings across individual studies will facilitate the development of a coherent theory or model of second language reading. This study attempted to: 1) identify the commonly used scoring/analysis systems in the L2 reading literature and 2) make an in-depth comparison among the scoring systems.

Literature Review

To examine the different methods researchers use to analyze the recall protocols, L2 reading studies published over the last twenty-five years (1980-2005) that used the recall task to measure reading comprehension were selected and compared. Six leading journals in the field of second language acquisition were surveyed: *Language Learning*, *Studies in Second*

Language Acquisition, Modern Language Journal, Applied Linguistics, TESOL Quarterly,

and *Foreign Language Annals*.   A total of 32 studies were identified.   The examination of

related research revealed that scoring methods used to analyze the recall protocols have varied.

Of the total 32 studies, 4 employed Meyer's (1975) scoring system (e.g., Connor, 1984; Allen

et al., 1988; Kim, 1995), 7 used Johnson's (1970) pausal unit system (e.g., Young, 1999;

Khaldieh, 2001; Chu et al. 2002; Maxim, 2002), and 19 employed the idea unit system (e.g.,

Carrell, 1984; Davis, Lange & Samuels, 1988; Barry & Lazarte, 1995; Horiba, 1996; Lee &

Riley, 1990; Ghaith & Harkouss, 2003).   Two studies, those by Donin and Silva (1993) and

Chen and Donin (1997), adopted Frederiksen's prepositional analysis, in which the logical

and semantic structures expressed in the text are examined.

Meyer's system, Johnson's system and idea unit system

Meyer's (1975) scoring system is based on case grammar and comprehensively reflects

both the structural characteristics and lexical units of a passage.   Procedures for the

development of a scoring template involve the following steps: 1) determination of the top

level structure; 2) determination of the macrostructural relationships by looking for lexical

items such as "because" and "therefore"; and 3) identification of lexical predicates and role

arguments.

Johnson's system (1970) is based on pausal units or breath groups.   The development of

a scoring template usually requires native speakers to read the passage aloud to themselves

and to mark all those places in the text where they paused.   Participants' recall protocols are

checked for the presence or absence of each pausal unit.   Research which has adopted

Johnson's system to analyze recall data differs in whether or not weighted scoring is used.

The use of weighed scoring requires native speakers to further divide all the pausal units into

four levels, with the lowest level being the least important 25 percent of the units in the

passage with each unit scored as 1 point, the next level being the next least important 25

percent of the units with each unit scored as 2 points and so forth.   Those who used the

unweighted scoring system simply treated each pausal unit equally and scored 1 point for

each unit.

Another scoring system is simply to count idea units that are recalled.   Unlike Meyer's

and Johnson's scoring systems, for which detailed guidelines on how to develop scoring

templates are provided in the literature, the identification of an idea unit when using the idea

unit system is relatively hard to do.   As Alderson (2000) pointed out, "An idea unit is

somewhat difficult to define, and this is rarely adequately addressed in the literature (p. 230)."

Chen and Donin (1997) also mentioned that idea unit analysis is "not only loose in its

psychological interpretation, but is often poorly defined and thus prone to subjectivity

(p.211)."   An examination of the selected L2 reading studies in which the analysis of idea

units was used revealed that the idea units have been identified differently across studies.

Some researchers adopted Bransford and Johnson's (1973) definition of the idea unit whereas

others divided the idea units based on the system developed by Bovair and Kieras (1985).

Still others simply mentioned the adoption of the idea unit system to analyze data without

explaining what constitutes an "idea unit."

Bransford and Johnson (1973) defined an idea unit as "corresponding either to individual sentences, basic semantic propositions, or phrases" (p.393).   This definition of idea units might, however, be interpreted differently.   Alderson (2000) illustrates how the identification of idea units might differ.   Take the following paragraph as an example.

> In free-recall tests (sometimes called immediate-recall tests), students are asked to read a text, to put it to one side, and then to write down everything they can remember from the text.   The free-recall test is an example of what Bachman and Palmer (1996) call an extended production response type (p.230).

According to Alderson (2000), this paragraph may be considered to contain 5 idea units: 1) Free-recall tests are sometimes called immediate-recall tests; 2) In free-recall tests, students read a text; 3) Students put the text to one side; 4) Students write down all they can remember; and 5) Bachman and Palmer (1966) call this test an extended production response type test.

The same paragraph can be considered to contain 15 idea units, if the researcher treats every content word or phrase as a separate idea (i.e., 1) free recall, 2) immediate recall, 3) tests, 4) students, 5) read, 6) one, 7) text, 8) put aside, 9) write, 10) all, 11) remember, 12) Bachman, 13) Palmer, 14) 1996, and 15) extended production response).

Unlike Bransford and Johnson's definition (1973), which might be subject to different interpretation, Bovair and Kieras (1985) provided clear rules on how to analyze the text into "idea/prepositional units" to develop scoring template for experimental purposes.   Bovair and Kieras's analysis system, developed from Kintsch (1974), also based on case grammar.

In order to facilitate the process of scoring the recall responses, they simplified the analysis procedures by disregarding tenses and auxiliaries in the texts and emphasizing the representation of contents.

Studies compare differences among scoring systems

While various scoring systems have been employed to analyze recall protocols, issue concerning whether different scoring systems bring about a different picture of L2 comprehension has not attracted much attention from researchers. An examination of the literature revealed that there existed only two studies directly examining the issue: Meyer (1985) and Bernhardt (1991). Meyer (1985) made a detailed comparison of the analysis systems of Kintsch (1974) and Meyer (1975). Since both systems provide specified hierarchies, Meyer first examined the differences in hierarchical structures derived from the two analysis approaches. The hierarchical structures of the Meyer approach were built on knowledge of logical relationships and discourse organizations, whereas the Kintsch approach focuses on word repetition. It was found that the correlation between the level of the propositions in hierarchy and subjects' recall performance (n=7) was higher in the Meyer approach (r=.50) than in Kintsch approach (r=-.16), which indicates that the Meyer approach is a better predictor of the data..

With regard to the differences in the scoring of Meyer and Kintsch approach, based on the performance of 9 subjects, the recall scores generated by both systems were highly correlated (r= .96). While the correlation between two analysis approaches was high, the

qualitative analysis revealed that Meyer's system is more satisfactory in studying the recall of less proficient readers and more sensitive in detecting the developmental differences in children.   In addition, Meyer's system allows researchers to score for the component relations and the content separately, which is not possible in Kintsch's system.

Although there are several strengths in Meyer's analysis system, the requirement of the expertise and time involved in the development of the scoring template and in scoring the recall protocols limits its potentials for being applied in large-scale assessment (Bernhardt, 1991).   In an attempt to search for an efficient scoring systems to increase the possibility of the use of the recall in nonresearch settings, Bernhardt (1991) conducted a validation study comparing the scores generated using Meyer's (1975) and Johnson's (1970) scoring systems.   Bernhardt (1991) used part of data in Allen et al. (1988) study, which had been scored using Meyer's system.   The recall protocols of 35 learners of German for two passages (i.e., one German newspaper article and one German business letter) were rescored using Johnson's system.   The results showed that the recall scores generated by both systems were highly correlated (i.e., r= .96 for the newspaper article and .85 for the business letter, respectively), which lent support to the use of Johnson's systems as an alternative for scoring the recall response, as advocated by Bernhardt (1991).

As aforementioned, available research that compares analyzed results employing different scoring/analysis systems is limited.   The nature of the variations in the analysis of the recall protocol warrants further investigation in order to make the comparability of

research findings possible.   This study examined whether and how variation in the analysis

of the recall protocol brings about different results.   The research questions this study

attempted to answer were: 1) Do different scoring systems bring about a different picture of

L2 comprehension? and 2) If so, how?

Methods

Participants

A total of 30 English major students participated in this study.   They were enrolled in

courses in the department of English at a university in central Taiwan and had been studying

English for at least nine years.   All subjects were native speakers of Mandarin Chinese

whose ages ranged from 19 to 21.

Data collection and analysis

One expository text introducing chemical substances was selected for this study.   The

reading passage, composed of 129 words (see Appendix I), was distributed.   Participants

were asked to read the text as many times as they needed and then to write down everything

they could remember from the text in their L1 when they felt ready.   They were encouraged

to recall as many details as they could and informed that the recall protocol task was not a

"main-idea-summarizing-exercise".

Since it is beyond the scope of this study to compare every scoring system mentioned in

the literature, this study focused on the comparison between the idea unit system developed

by Bovair & Kieras (1985)(idea units system I hereafter), the idea unit system illustrated by

Alderson (2000):counting content words or phrases as separate idea unit (idea units system II hereafter) and Johnson's pausal unit system.　To examine whether and how variation in the analysis of the recall protocol brings about different results, the participants' recall protocols were analyzed in terms of Johnson's unweighted pausal units system, idea unit system I and idea units system II.　Scoring templates were developed according to the procedures outlined by Johnson (1970), Bovair and Kieras (1985) and Alderson (2000).　To develop the scoring template following Johnson's (1970) pausal unit system, two native speakers of English was instructed to divide the reading text into pausal units based on normally paced oral reading (e.g., Some chemical substances/ have the potential/ to crystallize/ in two alternative ways). The pausal units identified by the two native speakers were quite similar except for the phrase "have the potential to crystallize."　One considered it as one unit whereas the other divided it into two units (i.e., "have the potential/ to crystallize").　In this case, the decision was made to select the narrower unit analysis.　Each pausal unit was listed, and participants' recall protocols were checked for the presence or absence of each unit.　The selected reading passage was then divided into 45 units using the pausal unit system, 47 units using idea unit system I and 66 units using idea unit system II.

　　After the marking scheme had been developed, I and one research assistant scored the recall protocol.　One point was given for each correctly recalled unit.　Scores were not given to those units recalled in English because in those cases there was some doubt as to whether readers actually understood the meaning of the units.　The inter-rater reliability

coefficient was found to be .86 for Johnson's system, .89 for idea unit system I and .87 for idea unit system II.

The proportion of idea units or pausal units recalled by each participant was calculated and then means and standard deviations of recall scores generated using different scoring approaches were calculated and compared. A repeated measures analysis of variance (ANOVA) was performed to detect whether the difference was statistically significant. In addition to the recall scores, an in-depth comparison was made unit by unit to reveal qualitative differences among scoring systems. During the process of scoring the recall responses using three different scoring systems, the researcher examined each student's recall protocol carefully and constantly compared the strength and weakness of each scoring system in evaluating the participant's recall responses.

Results

Tables 1, 2 and 3 illustrate the results of analysis employing the different analysis/scoring systems. As is shown in these tables, the nature of the divided unit varies because the principles underlying each analysis system differ. For example, the scoring template developed according to Johnson's pausal unit system rarely consisted of one-word units (e.g., only p40) because, as Bernhardt (1991) noted, "pausal unit endings are generally found at the end of a syntactically related unit" (p.209). Most of the units in the pausal unit system contain at least two words, that is, they are phrases or individual sentences. Unlike Johnson's system, the scoring template based on the idea unit system II, which counts every

content word as an idea unit, comprises mostly one-word units. Due to the fact that in Bovair and Kieras's system, the modifier is considered as one unit, the scoring scheme based on their approach, therefore, counts individual words, phrases and sentences as separate units. Each analysis system, though developed from different underlying principles, generates units of identical contents in the scoring template (e.g., p6 "for instance" in Johnson's system, p11 in Bovair & Kieras's system and p12 in idea unit system II; p17 "in diamonds"; p22 "This is why"; and p36 "which is why").

As shown in Tables 1, 2 and 3, the selected reading passage was divided into 45 units using the pausal unit system, 47 units using idea unit system I and 66 units using idea unit system II. Table 4 presents the mean proportions and standard deviations of recall scores generated using different scoring approaches. The results of the repeated measures of variance (ANOVA) for the mean scores among the three analysis systems are shown in Table 5. For an alpha level of .05, the results reveal that the difference was not statistically significant $F(2, 87)=0.47$, p>.05.

The correlational coefficients for the three approaches are presented in Table 6. The data revealed a highly positive correlation of .976 for the scores generated using Johnson's system and Bovair and Kieras's system, of .979 for the scores using Johnson's system and idea unit system II, and of 98.3 for the scores using Bovair and Kieras's system and idea unit system II.

While the scores generated using these analysis systems were highly correlated, detailed

comparisons of each scoring unit among scoring systems revealed qualitative differences in what the given score represents.   Scoring the recall protocols entails comparing the readers' responses with the scoring template.   It was found that the longer the unit is, the more difficult it is for researchers to determine the level of similarity in an objective way.   As can be seen in Tables 1, 2 and 3, units 1, 5, 9, 15, 20, 28, 37 and 45 of Johnson's system and units 7, 8, 13, 17, 25, and 31 of Bovair and Kieras's system contain more than one single "element" for the scorer to judge from.   Take unit 5 of Johnson's system as an example.   Unit 5 includes two elements --"Graphite and diamonds."   Of the total 30 students, 25 recalled "diamonds" correctly but only one knew what "graphite" is.   In scoring this unit, the researcher faced the dilemma of whether to give a point for the unit when the student only got half of it right.   To strictly follow the scoring template, the researcher chose not to give the partial credit for this case.

A similar problem occurred when scoring unit 37 "graphite feels slippery."   By following a strict scoring criterion, giving credit for the unit only if it was recalled verbatim or in a close paraphrase as recommended by Bovair and Kieras (1985), the scoring based on Johnson's system and Bovair and Kieras's system revealed that only one student recalled unit 37 correctly since only one student understood the word "graphite".   This analysis, however, fails to indicate that 11 out of 30 participants (37%) actually comprehended the meaning "feels slippery."   The same problem was also encountered in the scoring of units 1 "some chemical substances"; 8 "of pure carbon"; 15 "the carbon atoms"; 45 "with some other

chemical substances", etc., in which several modifiers are used to modify the head noun.

Aside from the difficulty of judging the accuracy of a multiple-element unit, the scorer also faces a dilemma of whether to repeatedly mark the same errors when scoring the recall protocol.   For example, in the assigned reading text, certain vocabulary items such as "graphite"(repeated 3 times), "diamond"(4 times), "crystallize"(2 times), "carbon"(4 times), "atom"(4 times), "chemical"(2 times), and "substance"(3 times) appeared more than once. Students indicated the word was unknown to them in their recall protocols either by spelling out the word in English or leaving a space.   With 7 words likely to affect 22 units out of a total of 66 for idea unit system II, 15 units out of 45 for Johnson's scoring system, and 15 units out of 47 for Bovair and Kieras' system, marking the same error repeatedly would result in a very different picture of learners' reading comprehension.

Another source of variation in scoring learners' recalls results from the difference among the scoring systems in the ability to signal the preservation of the syntactic structure of sentences as learners recalled it in their native language.   For example, unit 14 of Johnson's scoring system "in which" in the sentence "/The two substances/ differ from each other/only in the geometric pattern/ *in which*/ the carbon atoms/ are packed" is a syntactical marker in English indicating the use of a relative clause.   Since it is a common practice to have the students write the recall protocol in their L1 (Lee, 1986), the scoring of the protocols involves checking the participants' translation against each unit of the scoring template.   Unit 14 in Johnson's scoring system allows researchers to examine whether learners' recall used the

syntactic structure of the relative clauses.

Conclusion and Discussion

This study investigates whether and how different scoring systems evaluate or analyze second language readers' comprehension differently and is one of the few research studies that directly compares the scoring systems of the recall protocols. Although the principles behind each scoring system differ, the results revealed that the recall scores generated by the three systems are highly correlated (i.e. r= .97.6, .97.9 and .98.3).

These three scoring systems, however, differ in how straightforward and objective they are in allowing researchers to judge the correctness of learner's responses in their recalls. It appears that the longer the divided units are, the less likely it is for researchers to score the unit in a way that truly reflects learners' comprehension. As aforementioned, both Johnson's and Bovair and Kieras's scoring systems generate units containing multiple elements to judge from. Second language learners' recall protocols constantly manifest partial comprehension of the reading text due to their limited language proficiency. The scoring system with units containing multiple elements does not allow researchers to differentiate readers who partially comprehend a unit from those who do not understand any of it. In order to score the protocols in a way that truly reflects students' recall, the researcher may have to break the unit up into smaller units. The analysis based on idea unit system II, for example, provided a clearer picture of students' comprehension of the text, allowing the researcher to determine that 11 students (37%) actually understood the meaning of "feel slippery." The same

analysis system also allows researchers to score the unit 5 of Johnson system "Graphite and diamond" separately and, therefore, gain a better insight of what learners comprehended. One could also consider assigning more points for the longer units (i.e., the unit contains multiple elements) when the longer unit is not suitable for further division as in unit 45 "with some other chemical substances". For example, instead of assigning one point for the unit "with some other chemical substances", researchers can assign five points for this unit, which allows scorers to give a score distinguishing learners who recalled several modifiers besides head noun from those who did not.

As suggested by the evidence of this study, marking the same error repeatedly results in a very different picture of learners' reading comprehension. The examination of the extant literature, however, revealed that none of the study indicated how the researchers treated the repeat errors when scoring the recall protocols. In addition, very few studies mentioned whether the researcher followed a strict criterion in scoring the recall protocols. Provision of such details is essential in order to make possible the comparability and generalizability across studies.

Researchers need to make decision on how to treat the same errors prior to the scoring of the recall protocols according to their purposes of reading assessments. If the purpose of the test is to examine readers' understanding of the main idea of the passage, the selection of not marking the error repeatedly may be more efficient. On the other hand, if the purpose of the evaluation is to examine readers' comprehension of a text in detail, it would be better to

follow a strict scoring criteria and mark the same error repeatedly when it occurs.   Once the

decision is made, researchers need to be consistent through the process of the scoring in

treating the errors.

References

Alderson, J.C. (2000). *Assessing Reading*. Cambridge University Press.

Allen, E., Bernhardt, E., Berry, M., & Demel, M. (1988). Comprehension and text genre: An analysis of secondary school foreign language readers. *Modern Language Journal, 72*(2), 163-173.

Barry, S., & Lazarte, A. (1995). Embedded clause effects on recall: Does high prior knowledge of content domain overcome syntactic complexity in students of Spanish? *Modern Language Journal, 79*(4), 491-504.

Baylor, A., & McCormick, C. (2003). Thematic awareness and recall of information from text. *Reading and Writing, 16*, 247-258.

Bernhardt, E. (1991). *Reading development in a second language: Theoretical, empirical and classroom perspectives*.   Norwood, NJ: Ablex.

Bovair, S., & Kieras, D.E. (1985).   A guide to prepositional analysis for research on technical prose.   In B.K. Britton & J.B. Black (Eds.), *Understanding expository text: A theoretical and practical handbook for analyzing explanatory text* (pp. 315-362). Hillsdale, New Jersey: Erlbaum.

Bransford, J., & Johnson, M. (1973). Consideration of some problems of comprehension. W. Chase (Ed.), *Visual information processing* (pp. 383-483).   New York: Academic Press.

Brantmeier, C. (2003). Beyond linguistic knowledge: Individual differences in second language reading.   *Foreign Language Annals, 36*(1), 33-43.

Carrell, P.L. (1984). The effects of rhetorical organization on ESL readers.   *TESOL Quarterly, 19*, 441-69.

Chen, Q., & Donin, J. (1997). Discourse processing of first and second language biology texts: Effects of language proficiency and domain-specific knowledge.   *Modern Language Journal, 81*(2), 209-227.

Chu, H., Swaffer, J., & Charney, D. (2002). Cultural representations of rhetorical conventions: The effects on reading recall. *TESOL Quarterly, 36*(4), 511-541.

Connor, U. (1984). Recall of texts: Differences between first and second language learners. *TESOL Quarterly, 18*, 239-56.

Davis, J., Lange, D., & Samuels, J. (1988). Effects of text structure instruction on foreign language readers' recall of a scientific journal article.   *Journal of Reading Behavior,*

20(3), 203-214.

Davis, J.N. (1989). Facilitating effects of marginal glosses on foreign language reading. *Modern Language Journal, 73*(1), 41-48.

Donin, J., & Silva, M. (1993) The relationship between first-and second-language reading comprehension of occupation-specific texts. *Language Learning, 43*(3), 373-401.

Frederiksen, C.H. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology, 7*(31), 371-458.

Ghaith, G., & Harkouss, S. (2003). Role of text structure awareness in the recall of expository discourse. *Foreign Language Annals, 36*(1), 86-96.

Hammadou, J. (1991). Interrelationships among prior knowledge, inference, and language proficiency in foreign language reading. *Modern Language Journal, 75*(1), 27-38.

Horiba, Y. (1996). Comprehension processes in L2 reading. *Studies in Second Language Acquisition, 18*, 433-473.

Horiba, Y., van den Broek, & Fletcher, C. (1993). Second language readers' memory for narrative texts: Evidence for structure-preserving top-down processing. *Language Learning, 43*(3), 345-372.

Johnson, R. (1970). Recall of prose as a function of the structural importance of linguistic units. *Journal of Verbal Learning and Verbal Behavior, 9*, 12-20.

Khaldieh, S. (2001). The relationship between knowledge of Irabb, lexical knowledge and reading comprehension of nonnative readers of Arabic. *Modern Language Journal, 85*(3), 416-431.

Kim, S. (1995). Types and sources of problems in L2 reading: A qualitative analysis of the recall protocols by Korean high school EFL students. *Foreign Language Annals, 28*(1), 49-70.

Kintsch, W. (1974). *The representation of meaning in memory.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lee, J.F., & Riley, G.L. (1990). The effect of prereading, rhetorically-oriented frameworks on the recall of two structurally different expository texts. *Studies in Second Language Acquisition, 11*, 25-41.

Lund, R. (1991). A comparison of second language listening and reading comprehension. *Modern Language Journal, 75*(2), 196-204.

Maxim, H. (2002). A study into the feasibility and effects of reading extended authentic discourse in the beginning German language classroom. *Modern Language Journal, 86*(1), 20-35.

Mayer, R.E. (1985). Structural analysis of science prose: Can we increase problem-solving performance? In B.K. Britton & J.B. Black (Eds.), *Understanding expository text: A theoretical and practical handbook for analyzing explanative text* (pp. 65-87). Hillsdale, NJ: Lawrence Erlbaum Associates.

Meyer, B. (1975). *The organization of prose and its effects on memory.* Amsterdam: North-Holland.

Mohammed, M.A., & Swales, J.M. (1984). Factors affecting the successful reading of technical instructions. *Reading in a Foreign Language, 2*, 206-17.

Raymond, R. (1993). The effects of structure strategy training on the recall of expository prose for university students reading French as a second language. *Modern Language Journal, 77*(4), 445-458.

Sadoski, M., Goetz, E., & Avila, E. (1995). Concreteness effects in text recall: Dual coding or context availability? *Reading Research Quarterly, 30*(2), 278-288.

Sadoski, M., Goetz, E., & Rodriguez, M. (2000). Engaging texts: Effects of concreteness on comprehensibility, interest, and recall in four text types. *Journal of Educational Psychology, 92*(1), 85-95.

Young, D. (1999). Linguistic simplification of SL reading material: Effective instructional practice? *Modern Language Journal, 83*(3), 350-366.

Table 1. Analysis of the text using Johnson's (1970) pausal unit system

Units	frequency	Units	frequency
1. Some chemical substances	23 (77%)	24. so hard	22 (73%)
2. have the potential	25 (83%)	25. In graphite	1 (3.3%)
3. to crystallize	12 (40%)	26. the carbon atoms	7 (23%)
4. in two alternative ways	21 (70%)	27. are arranged	6 (50%)
5. Graphite and diamonds	1 (3.3%)	28. in flat hexagons	0
6. for instance	21 (70%)	29. layered on top of	12 (40%)
7. are both crystals	11 (37%)	30. each other	8 (27%)
8. of pure carbon	14 (47%)	31.The bonding	12 (40%)
9. Their atoms	6 (20%)	32. between layers	12 (40%)
10. are identical	11 (37%)	33. is weak	12 (40%)
11. The two substances	24 (80%)	34. and they therefore	9 (30%)
12. differ from each other	23 (77%)	35. slide over each other	7 (23.3%)
13.only in the geometric pattern	2 (7%)	36. which is why	13 (43%)
14. in which	5	37. graphite feels slippery	1 (3.3%)
15. the carbon atoms	10 (33%)	38. and is used	11 (37%)

16. are packed	14 (47%)	39. as a lubricant	0
17. In diamonds	18 (60%)	40. Unfortunately	16 (54%)
18. the carbon atoms	9 (30%)	41. You cannot crystallize diamonds	12 (40%)
19. are packed	15 (50%)	42. out of a solution	2 (7%)
20. in a tetrahedral pattern	0	43. by seeding them	1 (3.3%)
21. which is extremely stable	8 (27%)	44. as you can	5 (17%)
22. This is why	19 (67%)	45. with some other chemical substances	2(7%)
23. diamonds are	24 (80%)		

Table 2. Analysis of the text using Bovair and Kieras's (1985) idea unit system

Units	frequency	Units	frequency
P1(have substance the potential to P4)	25 (86%)	P25 (IsA diamonds hard)	22 (73%)
P2 (Mod substance **chemical**)	25 (86%)	P26 (In graphite)	1 (3.3%)
P3 (Mod chemical substance **some**)	22 (73%)	P27 (Arranged $ carbon atoms P28)	7 (23%)
P4 (crystallize substance)	12 (40%)	P28 (in hexagons P28)	0
P5 (In ways)	22 (73%)	P29 (Mod hexagons **flat**)	5 (17%)
P6 (Mod ways **two alternative**)	21 (70%)	P30 (Mod hexagons **layered on top of each other**)	8 (27%)
P7 (IsA graphite crystals P9 P10)	1 (3.3%)	P31 (IsA the bondings weak)	12 (40%)
P8 (IsA diamond crystals P9P10)	11(37%)	P32 (between layers)	12 (40%)
P9 (of crystal carbon)	14 (47%)	P33 (slide-over they each other)	7 (23%)
P10 (Mod carbon **pure**)	15 (50%)	P34 (IsA which why)	13 (43%)
P11 (for instance)	21 (70%)	P35 (feels graphite slippery)	11 (37%)
P12 (Possess two substances)	9 (30%)	P36 (used $ lubricant)	0
P13 (IsA atoms identical)	6 (20%)	P37 (Unfortunately)	16 (53%)
P14 (differ two substance)	23 (77%)	P38 (able you P40)	19 (63%)

P15 (differences are in pattern)	11 (37%)	P39 (negate p38)	19 (63%)
P16 (Mod patterns **geometric**)	2 (7%)	P40 (crystallize you diamonds)	12 (40%)
P17 (packed $ carbon atoms)	10 (33%)	P41 out of solution	2 (7%)
P18 (In diamonds)	18 (60%)	P42 by seeding them	1 (3.3%)
P19 (packed $ carbon atoms P18)	9 (30%)	P43 (able you P44)	5 (17%)
P20 (in a pattern P20)	10 (33%)	P44 (same as with substance)	5 (17%)
P21 (Mod pattern **tetrahedral**)	0	P45 (Mod p44 **other**)	12 (40%)
P22 (IsA P17 18 19 stable)	11 (37%)	P46(Mod other substances **some**)	2 (7%)
P23 (Mod stable **extremely**)	8 (27%)	P47 (Mod substance **chemical**)	11 (37%)
P24 (IsA this why P24)	19 (63%)		

Table 3. Analysis of the text using idea unit system II counting every content word as an idea unit

Units	frequency	**Units**	frequency
1.Some	23 (77%)	34. This is why	19 (67%)
2.chemical	25 (86%)	35. diamonds	24 (80%)
3.substances	25 (83%)	36. hard	22 (73%)
4.have the potential	25 (83%)	37. In graphite	1 (3.3%)
5.to crystallize	12 (40%)	38. carbon	7 (23%)
6.two	21 (70%)	39. atoms	9 (30%)
7. alternative	12 (40%)	40. are arranged	12 (40%)
8. ways	22 (73%)	41. flat	5 (17%)
9. Graphite	1 (3.3%)	42. hexagons	0
10. and	24 (80%)	43. layered	12 (40%)
11. diamonds	20 (67%)	44. on top of	13 (43%)
12. for instance	21 (70%)	45. each other	8 (27%)
13. crystals	11 (37%)	46. The bonding	12 (40%)
14. pure	15 (50%)	47. between layers	12 (40%)

15. carbon	14 (47%)	48. weak	12 (40%)
16. atoms	6 (20%)	49. slide over each other	7 (23%)
17. identical	11 (37%)	50. which is why	13 (43%)
18. two substances	24 (80%)	51. graphite	1(3.3%)
19. differ	23 (77%)	52. feels slippery	11 (37%)
20. only	8 (27%)	53. is used	11 (37%)
21. geometric	2 (7%)	54. as a lubricant	0
22. pattern	11 (37%)	55. Unfortunately	16 (53%)
23. carbon	10 (33%)	56. you cannot	19 (63%)
24. atoms	10 (33%)	57. crystallize	12 (40%)
25. are packed	14 (47%)	58. diamonds	17 (57%)
26.In diamonds	18 (60%)	59. out of a solution	2 (7%)
27.carbon	9 (30%)	60. by seeding them	1 (3.3%)
28.atoms	11 (37%)	61. as you can	5 (17%)
29.are packed	17 (57%)	62. with	5 (17%)
30. tetrahedral	0	63. some	2 (7%)
31. pattern	10 (33%)	64. other	12 (40%)
32. extremely	8 (27%)	65. chemical	11 (37%)
33. stable	11 (37%)	66. substances	14 (47%)

Table 4. Mean proportions and standard deviations generated using different scoring systems

Analysis method	n	Low	High	*M*	*SD*
Pausal units	30	.06	.72	.39	.25
Idea unit I	30	.02	.70	.38	.26
Idea unit II	30	.09	.71	.40	.24

Table 5. One-way ANOVA for the differences in the scores generated using different scoring systems

Source of variance	SS	*df*	MS	*F*
Between groups	.006	2	.003	.047
Within groups	5.466	87	.063	
Total	5.472	89		

$P > .05$

Table 6. Pearson Product-Moment Correlation Coefficients: Johnson's pausal unit system, Bovair and Kieras's idea unit system, and Idea unit system II (every content word).

	Johnson's	Bovair and Kieras's	Idea unit system II
Johnson's	1.000	.976***	.979**
Bovair and Kieras's		1.000	.983**
Idea unit system II			1.000

Appendix I

The reading text

Some chemical substances/ have the potential/ to crystallize/ in two alternative ways. Graphite and diamonds/, for instance,/ are both crystals/of pure carbon. /Their atoms/ are identical. /The two substances/ differ from each other/ only in the geometric pattern/ in which/ the carbon atoms/ are packed. /In diamonds/, the carbon atoms/ are packed/ in a tetrahedral pattern/ which is extremely stable. /This is why/ diamonds are/ so hard/. In graphite/ the carbon atoms/ are arranged/ in flat hexagons/ layered on top of/ each other/. The bonding/ between layers/ is weak/, and they therefore/ slide over each other/, which is why graphite/ feels slippery/ and is used/ as a lubricant/. Unfortunately/, you cannot crystallize diamonds/ out of a solution/ by seeding them/, as you can/ with some other/ chemical substances.