

## An Analysis of the Proposition Density, Sentence and Clause Types, and Non-Finite Verbal Usage in Two College Textbooks

Charmaine DeFrancesco and Kyle Perkins  
Florida International University, USA

**Abstract:** This study investigated the proposition density, sentence and clause type usage and non-finite verbal usage in two college textbooks. The teaching implications are presented.

A proposition is an idea unit; it is a statement that expresses a factual claim (Jay, 2003); it is the basic unit involved in the understanding and retention of text (Kintsch, 1974; Kintsch & Keenan, 1973). “Propositions correspond roughly to verbs, adjectives, adverbs, prepositions, and subordinating conjunctions (not nouns or pronouns)” (Covington, 2008, p. 2). Proposition density is an important factor in reading comprehension because of a proposition’s role in text comprehension and retention. In addition, “sentences in print often have a complex, embedded syntax that places demands on the reader’s working memory” (Graesser, McNamara, & Kulikowich, 2011, p. 226). The combination of text comprehension and retention, and demands on the reader’s working memory suggest that proposition density might be useful in the selection of college textbooks. The widely adopted readability formulas utilized in reading comprehension research do not estimate proposition density. Those readability formulas include Flesch-Kincaid Grade Level or Reading Ease (Klare, 1974), Degrees of Reading Power (Koslin, Zeno, & Koslin, 1987), Lexiles (Stenner, 2006), and Coh-Metrix (Graesser, McNamara, & Kulikowich, 2011). These single metrics are based on length of words and sentences (Flesch-Kincaid), readers’ performance on a cloze procedure (Degrees of Reading Power and Lexiles), and on various language-discourse levels (Coh-Metrix). Sentence and clause types and non-finite verbals are important in this research because they are directly related to complex, embedded syntax. Finally, the authors did not find any published studies of propositional density, sentence and clause type usage, and non-finite verbal usage in college textbooks. Therefore, the authors believe that the results of this study will contribute new information to the field and will establish baseline, or benchmark data for further comparative research on important factors that should be considered in the selection of textbooks for students who are reading to learn in post-secondary education.

### Purpose

The purpose of this empirical research study is to present a comparison of the proposition densities, the sentence and clause types, and the usage of non-finite verbals in two college textbooks. In this study, the authors define proposition density as the number of propositions in each sentence. The following example illustrates how the authors determined propositional density in this study:

In *The young gray squirrel has a very long tail*, there are five propositions identified below using the numbers one to five:

1. *has (squirrel, tail)*
2. *young (squirrel)*
3. *gray (squirrel)*
4. *long (tail)*

#### 5. *very* (<4>)

The proposition density of this sentence is 0.56 (5 propositions divided by the nine words contained in the sentence = 0.56). Furthermore, in numbers one to five above, 5. *very* (<4>) is an adverb that modifies the adjective *long* (*tail*) in 4. This is the preferred method for listing the propositions in a sentence.

### Method

The authors selected two widely known and widely adopted sports psychology textbooks to analyze: Weinberg and Gould's (2007) *Foundations of Sport and Exercise Psychology* and LeUnes' (2008) *Sport Psychology*. The authors selected *goal setting* as the common topic from both texts. Weinberg and Gould's text has an entire 20-page chapter on goal setting while the LeUnes' text contains only 52 sentences on goal setting; therefore, the authors randomly selected 52 sentences from Weinberg and Gould's text to match the content of LeUnes' sentences on goal setting.

The authors identified and recorded the number of propositions (i.e., verbs, adjectives, adverbs, prepositions, and subordinating conjunctions) in each of Weinberg and Gould's 52 sentences and the number of propositions in each of LeUnes' 52 sentences. The propositional density for each text was determined by dividing the total number of propositions identified by 52, the number of sentences examined. In addition, the researchers identified the number of different sentence types in the analyzed sentences (i.e., simple, compound, complex, and compound-complex), the number of clause types in the analyzed sentences (i.e., noun, adjectival, and adverbial), and the number of different non-finite verbals in the analyzed sentences (i.e., gerund, infinitive, and participle). Present and past participles were merged into one category—participle.

### Statistical Hypotheses

Because the authors could not find any studies in a review of the literature that presented data with which a comparison could be made, the authors developed the following hypotheses:

1. There is no significant difference in the average number of propositions per sentence in the two textbooks.

The alternative hypothesis is that there is a significant difference in the average number of propositions per sentence in the two textbooks.

2. There is no significant difference in the number of sentence types used in the two textbooks.

The alternative hypothesis is that there is a significant difference in the number of sentence types used in the two textbooks.

3. There is no significant difference in the number of clause types used in the two textbooks.

The alternative hypothesis is that there is a significant difference in the number of clause types used in the two textbooks.

4. There is no significant difference in the number of non-finite verbals used in the two textbooks.

The alternative hypothesis is that there is a significant difference in the number of non-finite verbals used in the two textbooks.

### Results

Table 1 displays the results of the proposition density analysis.

Table 1  
*Proposition Density*

Text	Weinberg and Gould	LeUnes
Number of propositions	635	648
Average number of propositions per sentence	12.21	12.46
Standard deviation	5.42	6.59
Range	4 - 26	3 - 29

$t = -0.20$ ,  $df$  102, n.s.

There were 635 propositions in the 52 sentences from the Weinberg and Gould text and 648 propositions in the 52 sentences from the LeUnes text. On average, there were 12.21 propositions in each Weinberg and Gould sentence and 12.46 propositions in each LeUnes sentence. A  $t$ -test for independent samples indicated that there was no significant difference between the means of the two samples of text; therefore, the data on which the test was based do not provide sufficient evidence to reject the null hypothesis.

The authors used the chi-square test for contingency tables for the analyses of sentence types, clause types, and non-finite verbal usage. Table 2 presents the data on sentence types.

Table 2  
*Sentence Types*

Text	Simple	Compound	Complex	Compound-complex	Total
Weinberg and Gould	27	1	20	4	52
LeUnes	32	4	14	2	52
Total	59	5	34	6	104

Chi-square = 3.94,  $df$  = 3, n.s.

The calculated chi-square of 3.94 is smaller than the tabled value for three degrees of freedom; therefore, both texts appear to use all four sentence types with equal frequency. The data on which the test is based do not provide sufficient evidence to reject the null hypothesis.

Table 3 presents the data for clause types in the samples of the two textbooks.

Table 3  
*Clause Types*

Text	Noun	Adjectival	Adverbial	Total
Weinberg and Gould	6	15	12	33
LeUnes	6	9	7	22
Total	12	24	19	55

Chi-square = 0.65,  $df$  = 2, n.s.

The calculated chi-square of 0.65 is smaller than the tabled value for two degrees of freedom; therefore, one can conclude that the two textbooks used noun, adjectival, and adverbial clauses with similar frequency. The data on which the test is based do not provide sufficient evidence to reject the null hypothesis.

Table 4 presents the data for the use of non-finite verbals found in the samples of the two texts.

Table 4  
*Non-finite Verbals*

Text	Gerunds	Infinitives	Participles	Total
Weinberg and Gould	25	23	18	66
LeUnes	41	8	12	61
Total	66	31	30	127

Chi-square = 12.16,  $df = 2$ ,  $p < .01$

The calculated chi-square of 12.16 exceeds the tabled value for two degrees of freedom at the  $p < .01$  level of significance. The data at hand are not compatible with the null hypothesis but are supportive of the hypothesis of a significant relationship between the text and the use of non-finite verbals. The test of independence does not specifically identify the difference between the two texts, but an examination of the bivariate frequency table suggests that LeUnes tends to use more gerunds than Weinberg and Gould, and that Weinberg and Gould use more infinitives than LeUnes.

### Discussion

The authors could find no criterion-referenced or norm-referenced data on propositional density, sentence and clause type frequency, and non-finite verbal usage to compare with the results of this research. That said, the authors are confident that both textbooks have many propositions embedded in each sentence, and further that the propositional density and complex, embedded syntax might create a cognitive overload for some readers. For example, compare the sentence that is offered to illustrate propositional density in the introduction of the paper, *The young gray squirrel has a very long tail*, which contains five propositions with the following sentence from the Weinberg and Gould (2007) textbook:

*In addition to improving our understanding of what makes goals more effective sport psychology researchers have also learned a good deal about the process of goal setting, including how people set goals, what goals are most important to people, what barriers impede goal attainment, and how different types of individuals differ in their goal setting.* (p. 349)

The authors are convinced that this sentence with its extremely dense propositional load and its complex, embedded syntax might create a cognitive overload for some readers for the following reasons. Learners can only attend to a finite amount of information at a given time due to the limited capacity of the working (short-term) memory. Incoming information from all the senses is stored in the sensory memory very briefly before it decays or is lost completely. The short-term store receives input from the sensory store and the long-term store, and this information is retained for approximately 30 seconds. Information is lost unless it is rehearsed; it is also lost if it is no longer needed. Information is transferred from the short-term store to the long-term store for fairly permanent storage (Atkinson & Shiffrin, 1968). If the amount of information to be processed exceeds a student's working memory capacity's to process it, then that student will have difficulty learning the material.

Miller (1956) advanced the notion that a person could hold from five to nine pieces of unrelated information in the short-term memory for processing, but more recent research indicates that the estimate should be lowered to as few as four (Cowan, 2001; Feldon, 2010; Janssen, Kirshner, Erkens, Krischner, & Pass, 2010). The Weinberg and Gould sentence shown above has over a dozen propositions in it. There is no metric to calculate how many pieces of

related and unrelated information a reader of that sentence would need to hold in short-term memory for eventual storage in the long-term memory, but the point should be obvious that this particular sentence entails a heavy processing load.

### Teaching Implications

The authors found that the two textbooks analyzed in this research contain, on average, a dozen propositions per sentence. The authors believe such an extremely dense propositional load might pose difficulties for some culturally and linguistically (CLD) students on their class rosters. Some CLD students have difficulty comprehending texts with a heavy propositional density because they may still be acquiring language and content concurrently. For this reason, the authors offer some suggestions based on their research findings to benefit CLD as well as majority students.

When a particular content objective is located in a very richly embedded propositional network, the instructor may want to focus on students' background knowledge of the content objective and supply the necessary vocabulary and background schemata needed to comprehend the passage. The instructor might also wish to provide steps to make the input more comprehensible. Comprehensible input includes (a) scaffolding, (b) breaking down the new concepts into smaller, more manageable parts, (c) advance organizers, (d) graphic organizers, (e) outlines of the materials to be covered, (f) semantic maps that show the relationships of the concepts or knowledge in the text, and (g) verbal scaffolding (Echevarria, Short, & Peterson, 2012). Graphic organizers are especially useful for depicting relationships between nomenclature and ideas within a content objective and a learning task and to reduce a novel content objective into small, more manageable parts. Instructors could also use (a) thematic maps to depict hierarchical relationships, (b) network trees to illustrate superordinate or subordinate elements, and (c) spider maps to relate non-hierarchical information to a topic sentence or to a thesis statement.

### Summary

The authors have presented data on the average number of propositions per sentence and the usage of sentence types, clause types, and non-finite verbals in two college textbooks. There are no comparable data for comparison. The authors have also related the findings to factors that are known to influence text comprehension and retention and to short- and long-term memory retention. The authors concluded the paper with some pedagogical suggestions based on the results of the analyses. The data presented in this paper provide a benchmark for further cross-validation studies with other texts and for cross-validation, concurrent validation studies with the most common readability formulas used in reading comprehension research.

### References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*, (Vol. 2, pp. 89-195). New York, NY: Academic Press.
- Covington, M. A. (2008). Idea density – A potentially informative characteristic of retrieved documents. Retrieved from <http://www.ai.uga.edu/caspr.Covington-2009-Idea-Density-paper-SEC09-060.pdf>
- Cowan, N. (2001). The magical number four in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87-185.
- Echevarria, J., Short, D. J., & Peterson, C. (2012). *Using the SIOP model with pre-k and kindergarten English learners*. Boston, MA: Pearson.
- Feldon, D. F. (2010). Why magic bullets don't work. *Change*, 42(1), 5-21.

- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multi-level analyses of text characteristics. *Educational Researcher, 40*, 223-234.
- Janssen, J., Kirshner, F., Erkens, G., Kirschner, P. A., & Paas, F. (2010). Making the black box of collaborative learning transparent: Combining process-oriented and cognitive load theory. *Educational Psychological Review, 22*, 139-154.
- Jay, T. B. (2003). *The psychology of language*. Upper Saddle River, NJ: Pearson.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.
- Kintsch, W., & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology, 5*(3), 257-274.
- Klare, G. R. (1974-1975). Assessing readability. *Reading Research Quarterly, 10*, 62-102.
- Koslin, B. I., Zeno, S., & Koslin, S. (1987). *The DRP: An effective measure in reading*. New York, NY: College Entrance Examination Board.
- LeUnes, A. (2008). *Sport psychology* (4<sup>th</sup> ed.). New York, NY: Psychology Press.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63*, 81-97.
- Stenner, A. J. (2006). Measuring reading comprehension with the Lexile framework. Durham, NC: Metametrics.
- Weinberg, R. S., & Gould, D. (2007). *Foundations of sport and exercise psychology* (4<sup>th</sup> ed.). Champaign, IL: Human Kinetics.