# Reading into the Text: Investigating the Influence of Text Complexity on Cognitive Engagement


Ben Vega
University of Memphis
365 Innovation Drive
Memphis, TN 38152
bavega@memphis.edu

Shi Feng
University of Memphis
365 Innovation Drive
Memphis, TN 38152
sfeng@memphis.edu

Blair Lehman
University of Memphis
365 Innovation Drive
Memphis, TN 38152
balehman@memphis.edu

Arthur C. Graesser
University of Memphis
365 Innovation Drive
Memphis, TN 38152
graesser@memphis.edu

Sidney D'Mello
University of Notre Dame
384 Fitzpatrick Hall
Notre Dame, IN 46556
sdmello@nd.edu



## ABSTRACT

Boredom and disengagement have been found to negatively impact learning. Therefore, it is important for learning environments to be able to track when students disengage from a learning task. We investigated a method to track engagement during self-paced reading by analyzing reading times. We propose that a breakdown in the relationship between reading time and text complexity can reveal disengagement. A discrepancy (or decoupling) between attention resources and text complexity was computed via the absolute difference between reading times and the text's Flesch-Kincaid Grade Level, a measure of text complexity. As expected, decoupling varied as a function of text complexity. We also found that text complexity differentially impacted decoupling profiles for different types of participants (i.e., high vs. low comprehenders, fast vs. slow readers). These results suggest that decoupling scores may be a viable method to track disengagement during reading and could be used to trigger interventions to help students re-engage with the text and ultimately learn the material more effectively.



## Keywords

Engagement, boredom, reading, text complexity


## 1. INTRODUCTION

It is widely acknowledged that engagement in a learning task is a necessary (but not sufficient) condition to achieve learning gains. There is also data to support this assumption. For example, student engagement was found to positively correlate with learning during interactions with an intelligent tutoring system (ITS) called AutoTutor, whereas boredom negatively correlated with learning [1]. Given this relationship, learning environments should seek to maximize engagement and minimize boredom and disengagement.

A variety of methods have been used to track student engagement during learning. These include body movements, facial expressions, aspects of language and discourse, self-reports, and observations by trained judges [2-4]. While these measures focus on the *affective* dimension of engagement, here we propose a method to track *cognitive* engagement during a reading task [5]. We posit that student engagement can be measured and tracked through a comparison of reading times and text complexity. The dance between reading times and text complexity can either align (e.g., the text becomes more difficult and reading times increase) or misalign (e.g., the text becomes more difficult but reading times do not reflect that). When discrepancies between reading times and text complexity occur, it may be indicative of students disengaging from the current learning task because they are not appropriately allocating resources to meet task demands.

Our work is grounded in previous research that has shown that reading times are robustly predicted by such language and text characteristics as word length and frequency, sentence length, and other discourse characteristics [6-7]. However, there is a lack of research that uses reading time measures to assess engagement in reading at a fine grained level. In the present paper we investigate the relationship between reading times and text complexity as assessed via Flesch-Kincaid Grade Level scores [8]. The present study investigates the discrepancy (or decoupling) between reading times and text complexity as a new method to track student engagement during a self-paced reading task.

## 2. METHOD

### 2.1 Participants

There were 64 participants in the present study who were recruited from Amazon's Mechanical Turk™ (AMT). AMT acts as a mediator between researchers and individuals to allow people to complete psychological tasks online for monetary compensation. Participants were limited to native English speakers of at least 18 years of age. On average, it took participants 33 minutes to complete the study and they were compensated US $4 for their participation. Past research suggests AMT is a reliable and valid source for collecting experimental data [9].

### 2.2 Materials

### 2.2.1 *Texts*

The texts were adapted from the electronic textbook that accompanies the *Operation ARA!* ITS with conversational agents [10]. ARA helps students learn about research methodology through electronic texts, conversations with agents, and critiquing flawed science. Each text discussed one of four research methods topics: causal claims, dependent variable, experimenter bias, and replication. A text began with a real world situation to ground the research methods concept being discussed. The text then continued with explanations and examples that suggest more generalized uses for the concept. Each text was approximately 1500 words long. Order of texts was counterbalanced across participants with a Latin Square.

### 2.2.2 *Knowledge Assessment*

Assessment of research methods knowledge was conducted after participants read each of the four texts. Each assessment consisted of six multiple-choice questions pertaining to the research methods concept in the text. The questions were developed using the Graesser-Person question asking taxonomy [11] specifically targeting logical, causal, or goal-oriented reasoning.

## 2.3 Procedure

Participants signed an electronic consent form and read the instructions for the self-paced reading task. Self-paced reading was adopted for this task to eliminate any pressures from time constraints. Participants were then presented with the first of four texts. A sentence-by-sentence reading paradigm was used in which texts were presented one sentence at a time and participants pressed the space bar to move on to the next sentence. Reading times were collected for each individual sentence from each of the four texts. After participants read the first text, they were presented with the knowledge assessment for the research methods concept in that text. Participants then began the second text and repeated this pattern for all four texts.

# 3. RESULTS & DISCUSSION

The analyses are divided into two sections. First, we discuss how the decoupling score was computed. Second, we explore the relationship between the decoupling score and text complexity.

## 3.1 Decoupling Score

The decoupling score was computed as a measure of the degree to which participants were appropriately allocating their attention based on text characteristics. In other words, as the text became more difficult, did participants spend more time reading the text? To compute this score each text was divided into overlapping groups of three sentences (triplets) such that sentences 1, 2, and 3 were one triplet, sentences 2, 3, and 4 were a second triplet and so on.

For each triplet, two values were used to compute the decoupling score. First, the total reading times for the three sentences in a triplet were summed and then standardized (i.e., converted to a z-score) on the subject level. Second, the Flesch-Kincaid Grade Level (FKGL) was computed for each triplet. The FKGL ranges from grades 1-12 and assesses the difficulty of a text based primarily on sentence length and the number of syllables. The FKGL was then also standardized (i.e., converted to a z-score) based on the TASA corpus, as computed by the Coh-Metrix text analysis tool [12].

The decoupling score was computed using the standardized reading time and FKGL scores for each triplet by taking the absolute difference between the two scores. Thus, in the present analysis, we are only focusing on the magnitude of decoupling and not the direction of decoupling (i.e., allocating too much or too little time based on text complexity).

## 3.2 Decoupling & Text Complexity

We investigated how decoupling scores varied as a function of text complexity, as assessed by FKGL. To investigate this relationship we plotted the decoupling score (y-axis) as a function of the standardized FKGL (x-axis) (see Empirical in Figure 1). FKGL scores were divided into ten equal partitions and the average is plotted in Figure 1. It is possible that the observed curve in Figure 1 is an artifact of the computations used to create the decoupling score. To address this issue, we constructed a randomly shuffled surrogate of the corpus. In this surrogate corpus, the FKGL score for each triplet was preserved, however, the ordering of the reading times was randomized. Ten surrogate corpora were constructed for each participant. Decoupling scores were then computed for each surrogate corpus and the average was used for the Shuffled curve in Figure 1.

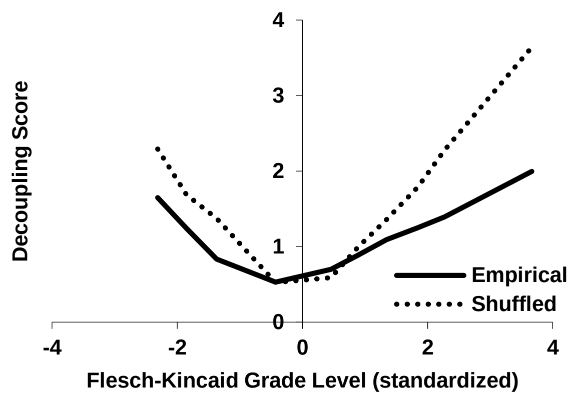


**Figure 1. Decoupling scores as a function of text complexity**

As can be seen in Figure 1, the curvilinear shape does partially appear to be an artifact of the computations used in the present analyses. However, the two curves were not identical. We investigated the differences between the two curves by conducting a 2 (curve: empirical or shuffled) x 10 (FKGL partition) repeated measures ANOVA. The ANOVA revealed that there were significant main effects comparing the two curves [$F(1,63) = 4.55$, $p < .001$, $\eta^2 = .343$] and the 10 partitions [$F$(9,567 = 715, $p < .001$, $\eta^2 = .919$] as well as a significant curve × partition interaction [$F(9,567) = 3.03$, $p < .001$, $\eta^2 = .639$]. Post hoc analyses with Bonferroni correction showed that the empirical curve and shuffled curve significantly differed at all partitions, suggesting that the empirical curve does differ from chance.

A closer examination of the empirical curve shows that when text complexity is near the mean level of complexity, i.e. FKGL (standardized) = 0, decoupling is less than when compared to the extremes of text difficulty (i.e., very easy or very hard). One way of interpreting this pattern is that the participants read at a pace appropriate for sentences with average complexity, but failed to adjust their reading speed in accordance with the more extreme levels of text complexity. This would result in the participants either spending too much or too little time on the easier and the harder portions of the texts.

It could be the case, however, that the relationship between text complexity and decoupling is obscured when all participants are combined into one group. To investigate this potential issue, we divided participants based on reading speed (fast, slow) and comprehension (i.e., score on knowledge assessment; high, low). Participants were divided via a median split for reading speed and comprehension, resulting in four groups: fast reader-high comprehender (FR-HC, $N$ = 16), fast reader-low comprehender (FR-LC, $N$ = 17), slow reader-high comprehender (SR-HC, $N$ = 19), and slow reader-low comprehender (SR-LC, $N$ = 12). Scores for the 4 groups are plotted in Figures 2-5.

ANOVAs showed that there were significant main effects and interaction terms for all four groups (*p's* < .05), with the exception that the curve main effect for the FR-LC group was not significant ($p$ = .973). These poor readers were essentially insensitive to text complexity, as would be expected.

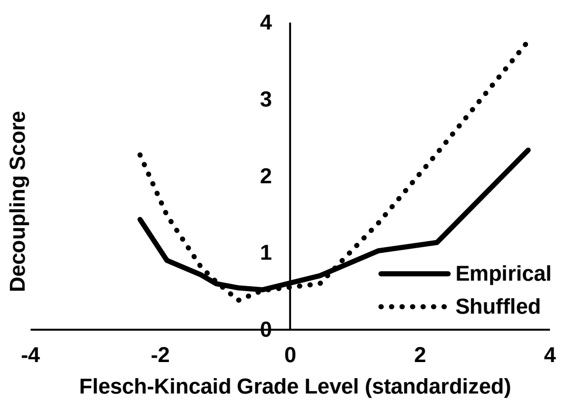


**Figure 2. Decoupling scores as a function of text complexity for the SR-HC group**

The SR-HC group was most sensitive to the more complex portions of the text. That is when the text was more difficult, this group had less decoupling than chance (i.e., Shuffled curve in Figure 2). On the other hand, the FR-LC group did not vary their reading behavior based on the text complexity. This can be seen in the close proximity of the Empirical and Shuffled curves in Figure 3.

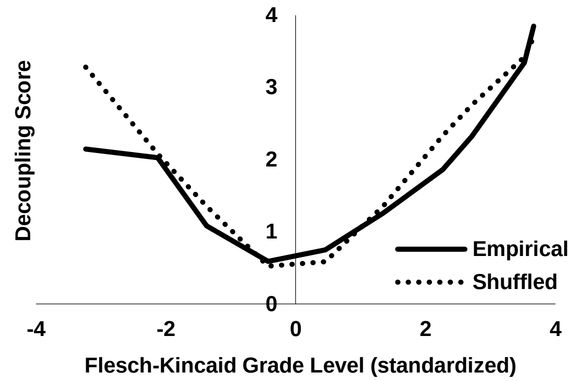


**Figure 3. Decoupling scores as a function of text complexity for the FR-LC group**

The FR-HC showed greater decoupling than chance (see Figure 4) at almost all levels of text complexity. This suggests that participants in this group could have been extremely vigilant to the text complexity, spending much less time on easy texts and much more time on difficult text segments. However, it is difficult from the present data to determine why this group of participants had these decoupling patterns.

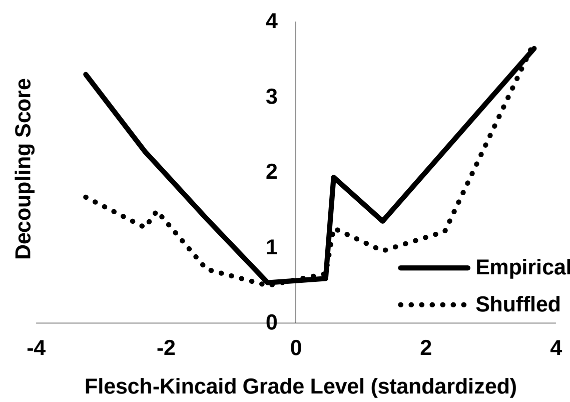


**Figure 4. Decoupling scores as a function of text complexity for the FR-HC group**

The SR-LC group had less than expected decoupling at extreme levels (easy or difficult) (see Figure 5). In other words, the more the text complexity deviated from the mean in either direction the decoupling was less than expected. This pattern may indicate that participants in this group needed the complexity level of the text to be more explicit or obvious for them to adapt their reading behavior. However, the overall decoupling score for this group did increase at the extremes, although this was still less than chance (see Shuffled curve in Figure 5). Unfortunately, it is somewhat difficult to interpret these results without knowing the direction of decoupling.

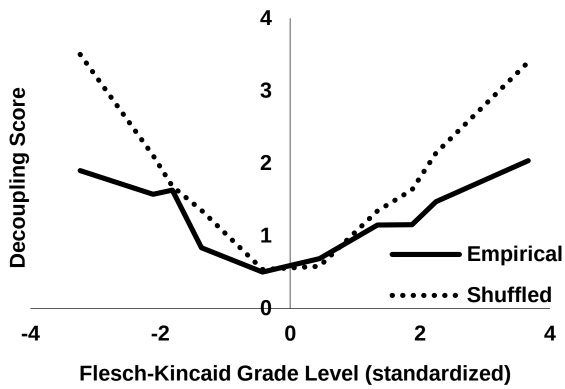


**Figure 5. Decoupling scores as a function of text complexity for the SR-LC group**

# 4. CONCLUSION

This study investigated a new method of measuring and tracking cognitive engagement during reading. The decoupling score was derived from the absolute difference between reading times and text complexity. We propose that this measure assesses cognitive engagement because if readers are engaged with the text, then their reading times should be adjusted based on text complexity. In other words, as the text becomes easier, reading times should become relatively faster and, conversely, as the text becomes more difficult reading times should become relatively slower. We found evidence that the relationship between reading time and text complexity did seem to reveal patterns of disengagement. Moreover, we found that the relationship between decoupling and the complexity of the text varies based on individual differences in reading speed and comprehension.

Despite these promising initial findings, we were not able to completely explain the patterns of decoupling for all types of participants. In particular, the relationship between decoupling pattern and comprehension scores was not clearly revealed in the differences between the empirical data and the shuffled surrogate corpus for participants classified as fast reader-high comprehenders. This highlights a limitation of using Flesch-Kincaid Grade Level to assess text complexity. Flesch-Kincaid assesses text complexity at a rather shallow level. It may be the case that more nuanced measures of text complexity will be able to shed more light on how decoupling impacts comprehension. Thus, in future work we plan to investigate more differentiated measures of text complexity, such as narrativity, syntactic simplicity, referential cohesion, word concreteness, and situation model cohesion using Coh-Metrix [12]. We are specifically targeting cohesion because past research has shown that cohesion and breakdowns in cohesion impact learning as well as interact with prior knowledge [13].

Student engagement over the course of a learning experience is a vital issue. This paper provides insight on how text complexity can factor into cognitive engagement levels and a possible measure for it. More importantly, this measure may be capable of tracking students' cognitive engagement across a span of text by simply using reading times and text characteristics (e.g., complexity). This measure of cognitive engagement could then be used to create texts that adapt in complexity level to increase cognitive engagement and maximize learning.

# 5. REFRENCES


[1] D'Mello, S. & Graesser, A. 2012. Dynamics of affective states during complex learning. *Learning and Instruction*, 22, 145-157.

[2] D'Mello, S., & Graesser, A. 2010. Multimodal semi-automated affect detections from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, 20, 147-187.

[3] Jackson, S., & Marsh, H. 1996. Development and validation of a scale to measure optimal experience. The Flow State Scale. *J. of Sport & Exercise Psychology*, 18, 17-35.

[4] Baker, R., D'Mello, S., Rodrigo, M., & Graesser, A. 2010. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *Int. J. of Human-Computer Studies*, 68, 223-241.

[5] Fredricks, J., Blumenfeld, P., & Paris, A. 2004. School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74, 59-109.

[6] Haberlandt, K., & Graesser, A. 1985. Component processes in text comprehension and some of their interactions. *J. of Experimental Psychology: General*, 114, 357-374.

[7] Graesser, A., & McNamara, D. 2011. Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3, 371-398.

[8] Klare, G. 1974. Assessing readability. *Reading Research Quarterly*, 10, 62-102.

[9] Mason, W., & Suri, S. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44, 1-23.

[10] Halpern, D., Millis, K., Graesser, A., Butler, H., Forsyth, C., & Cai, Z. 2012. Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity*, 7, 93-100.

[11] Graesser, A., & Person, N. 1994. Question asking during tutoring. *American Educational Research Journal*, 31, 104-137.

[12] Graesser, A., McNamara, D., & Kulikowich, J. 2011. Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40, 223-234.

[13] Ozuro, Y. Dempsey, K., McNamara, D. 2009. Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, 19, 228-242.