

Text-based recall and extra-textual generations resulting from simplified and authentic texts

Scott A. Crossley
Georgia State University
United States

Danielle S. McNamara
Arizona State University
United States

Abstract

This study uses a moving windows self-paced reading task to assess text comprehension of beginning and intermediate-level simplified texts and authentic texts by L2 learners engaged in a text-retelling task. Linear mixed effects (LME) models revealed statistically significant main effects for reading proficiency and text level on the number of text-based propositions recalled: More proficient readers recalled more propositions. However, text level was a stronger predictor of propositional recall than reading proficiency. LME models also revealed main effects for language proficiency and text level on the number of extra-textual propositions produced. Text level, however, emerged as a stronger predictor than language proficiency. Post-hoc analyses indicated that there were more irrelevant elaborations for authentic texts and intermediate and authentic texts led to a greater number of relevant elaborations compared to beginning texts.

Keywords: text readability, text comprehension, L2 reading, text simplification

Second language (L2) readers and teachers generally have two choices when selecting reading texts: authentic texts that were developed for first language (L1) readers or texts that have been linguistically simplified to increase comprehension. There are obvious trade-offs between the two choices and neither is optimal. For instance, authentic texts, while preserving natural language complexity and cultural relevance, are often difficult to process and comprehend because of their use of lexically sophisticated words and chunks, syntactic complexity, and lack of explicit cohesive devices (Crossley, Allen, & McNamara, 2011, 2012; Crossley, Louwrese, McCarthy, & McNamara, 2007; Crossley & McNamara, 2008). Simplified texts, on the other hand, appear to be easier to process and comprehend because of the manipulation of linguistic features (Crossley, Yang, & McNamara, 2014; Long & Ross, 1993; Oh, 2001; Tweissi, 1998; Yano, Long, & Ross, 1994), but the process of simplification can rob the texts of their natural rhythm and cultural significance (Little, Devitt, & Singleton, 1989; Long & Ross, 1993).

One limitation on the previous behavioral studies conducted on text simplification and its relation to text comprehension has been the manner in which comprehension has been defined. The majority of previous studies have measured comprehension through comprehension questions (e.g., true/false or multiple choice questions; Crossley et al., 2014; Long & Ross, 1993; Oh, 2001; Tweissi, 1998; Yano, Long, & Ross, 1994). While comprehension questions provide an indication of comprehension, they have limitations. These limitations include the notions that comprehension questions generally query only a small number of the ideas found in a text, can be correctly guessed (Day & Park, 2005), and do not reflect theoretical assumptions guided by comprehension models (Kintsch, 1988, 1998).

In this study, we examine text comprehension at various text levels (authentic texts and texts simplified to the beginning and intermediate levels) using a text-retelling paradigm that is embedded in a self-paced reading experiment. Text retelling allows readers to freely produce the propositions they recall from reading the text as well as extra-textual elaborations. The number of propositions recalled is limited only by the time available for the retelling, and retellings by their very nature do not allow readers to guess at their answer. In addition, the use of propositions as a measure of comprehension is firmly rooted in a number of theoretical and empirical accounts of reading (e.g., the Construction-Integration model of comprehension; Kintsch, 1988, 1998). Thus, our goal in this study is to assess the relations between text simplification and comprehension in L2 readers using propositional data. In addition, we examine additional non-textual factors that are often strongly related to reading comprehension including reader background knowledge, reading proficiency, and overall L2 language proficiency (Crossley et al., 2014). These factors have been generally neglected in previous research on the effects of text simplification (e.g., Long & Ross, 1993; Oh, 2001; Tweissi, 1998; Yano et al., 1994). Such an approach allows us to answer the following research questions:

1. Are there differences in text comprehension as measured by propositions recalled for L2 readers among texts simplified to the beginning and intermediate level and authentic texts?
2. Does an L2 reader's background knowledge, language proficiency, or reading proficiency aid in text comprehension?
3. Do texts simplified to the beginning and intermediate levels lead to a greater or smaller number of extra-textual propositions produced as compared to authentic texts?
4. Does an L2 reader's background knowledge, language proficiency, or reading proficiency lead to a greater or smaller number of extra-textual propositions?

Text Simplification

Authentic texts are unmodified texts that were originally created to fulfill a social purpose in a first language community (Little, Devitt, & Singleton, 1989). Often authentic texts are modified to make them more linguistically accessible for L2 readers. In this way, material developers hope to maintain the cultural relevance of the text while, at the same time, simplifying the text to make it more comprehensible. Such text modifications generally occur at the syntactic and lexical level (Hill, 1997), but modifications are also common at the level of cohesion (Crossley, Louwse, McCarthy, & McNamara, 2007; Crossley & McNamara, 2008). Some authentic texts are also

simplified through elaboration in order to clarify the content of the text and simplify the text structure through the repetition of key ideas and the paraphrasing of difficult terms (Yano, Long, & Ross, 1994), although such elaboration appears to lead to decreased readability (Long & Ross, 1993; McNamara, Kintsch, Songer, & Kintsch, 1996).

While there are many approaches to text simplification, such as adapting or abridging original texts and writing texts specifically to practice a grammar or linguistic form, all simplified texts share the same goal: reducing the cognitive load and increasing text comprehensibility on the part of the L2 reader. When simplifying a text, material developers generally follow two approaches: a structural approach or an intuitive approach (Allen, 2009). In an intuitive approach, authors use their experiences as a language teacher, language learner, and/or materials developer to guide them in the process of text simplification. Thus, an intuitive approach relies on an author's subjective judgment of what learners at a particular level are able to comprehend and read (Allen, 2009). A structural approach to simplification relies on authors using pre-defined word and structure lists. These approaches are most commonly used in graded reader texts that are linked to practices of extensive reading. In a similar fashion, authors may rely on traditional readability formulas that assess text readability based on sentence length and word length to simplify text. While such readability formulas can be successful at predicting L1 text readability, they are widely criticized as weak indicators of comprehensibility (Carrell, 1987; Crossley, Greenfield, & McNamara, 2008; Davison & Kantor, 1982). Of these two approaches to text simplification (intuitive and structural), intuitive approaches are more common (Crossley, Allen, & McNamara, 2012; Simensen, 1987).

Simplification and Textual Effects

The reasons behind text simplification are clearly defined. However, the linguistic effects of such modifications on texts were unclear until recently. That is to say, material developers routinely simplified texts in order to make them more readable and comprehensible, but to what degree these modifications led to linguistic differences as compared to authentic texts remained uncertain. In a series of studies conducted by Crossley and colleagues, the linguistic differences between authentic and simplified texts (Crossley et al., 2007; Crossley & McNamara, 2008) and between levels of simplified text (Crossley, Allen, & McNamara, 2011, 2012) were clarified.

These studies generally supported the notion that the process of text simplification led to significant changes in the linguistic structure of texts, both when comparing simplified to authentic texts and when comparing levels of simplified texts. The findings provided evidence that simplification should lead to texts being easier to read and comprehend. For example, Crossley et al. (2007) and Crossley and McNamara (2008) reported that authentic texts used for beginning and intermediate L2 learners were syntactically more complex, contained a greater density of logical connectors, contained greater lexical sophistication (e.g., more infrequent words, less specific words, words with more senses, and less familiar words) and had lower levels of cohesion (e.g., less lexical co-reference and semantic overlap) than simplified texts used at the same levels. In reference to texts simplified to specific levels (i.e., text simplified for advanced, intermediate, and beginning level L2 readers), Crossley et al. (2012) found that advanced level simplified texts when compared to beginning simplified texts were more complex

lexically (e.g., contain greater lexical diversity, more infrequent words, more unfamiliar words, and less concrete words), syntactically (e.g., have less syntactic similarity and more words before the main verb), and cohesively (e.g., less given information, less semantic co-referentiality, and less noun overlap). These studies indicate that the process of simplification leads to the creation of texts that should be easier to process and comprehend for L2 readers.

Simplification and Text Comprehension

While linguistic differences between simplified and authentic texts and differences between texts simplified to various proficiency levels are indicative of potential processing differences, they do not provide evidence of processing differences. For that, behavioral studies are needed. Those behavioral studies that have examined the effects of text simplification on L2 readers have generally supported the notion that simplified texts do lead to both faster reading times and improved text comprehension. For instance, Yano et al. (1994) reported that simplified texts as compared to authentic texts, increased text comprehension. In more recent studies, Tweissi (1998) and (Oh, 2001) also reported that simplification positively affected L2 students' overall reading comprehension. However, at least one study (Long & Ross, 1993) indicates complications with text simplification that raise cautionary notes. Similar to other studies, Long and Ross (1993) reported that texts linguistically simplified using traditional readability formulas led to greater comprehension in L2 readers when compared to authentic texts. However, Long and Ross also reported that readers' English proficiency level and reading comprehension scores affected text comprehension with higher proficiency learners and readers exhibiting better text comprehension.

While these studies collectively support the use of simplified over authentic texts in terms of text comprehension, potential limitations in their experimental designs indicate that the results should be interpreted with caution. For instance, the Tweissi (1998) study did not statistically control for potential linguistic differences between text conditions and Long and Ross (1993) and Yano et al. (1994) relied solely on traditional readability formulas, which are limited in the number of linguistic features they measure, to assess differences between simplified and authentic text. More importantly, many of the studies did not control for reading proficiency (Long & Ross, 1993; Oh, 2001; Tweissi, 1998; Yano et al., 1994), language proficiency (Yano et al., 1994; Tweissi, 1998), or background knowledge (Long & Ross, 1993; Oh, 2001; Tweissi, 1998; Yano et al., 1994) when assessing text comprehension. Reading and language proficiency (Buswell, 1922) along with background knowledge are important predictors of readability and text comprehension (McNamara et al., 1996; Shapiro, 2004).

To at least partially address these limitations, Crossley et al. (2014) used a moving windows self-paced reading task to examine differences in reading times and comprehension for L2 learners reading authentic texts and texts simplified to the beginning and intermediate levels. In addition to controlling for linguistic differences in the text using the computational tool Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004; McNamara & Graesser, 2012; McNamara, Graesser, McCarthy, & Cai, 2014), Crossley et al. also controlled for the reading proficiency, language proficiency, and background knowledge of the L2 participants. Crossley et al. used a moving windows self-paced reading task in order to simulate eye movement data (Just, Carpenter, & Woolley, 1982).

Crossley et al. (2014) found that beginning level texts were processed faster and were more comprehensible than intermediate level and authentic texts. The effect of text type on comprehension remained significant within an analysis of covariance controlling for language proficiency (i.e., TOEFL scores), reading proficiency (i.e., Gates-MacGinitie scores), and background knowledge, but not for reading times. However, the results also indicated that text simplification may be beneficial only if the L2 reader does not have strong background knowledge of the topic and that the use of simplified texts is more beneficial to beginning readers than advance readers. In addition, while text simplification appears to decrease reading times, reading ability is likely a stronger predictor of reading time. Thus, the effects of text simplification are moderated by the individual differences of the reader.

Propositional Approaches to Investigating Text Comprehension

A limitation of the studies discussed thus far is their reliance on comprehension questions (i.e., true/false or multiple choice questions) as a marker of text comprehension. Answering such questions relies on recognizing explicit text. However, in a number of network-based models of comprehension, comprehension is estimated by the quality of the reader's mental representation of the information in the text and meaning is represented in terms of propositions. One such model is the construction-integration model (Kintsch, 1988, 1998; van Dijk & Kintsch, 1983). Accordingly, a proposition is the smallest unit of meaning that can be represented in a predicated-argument form, represents one complete idea, and contains a truth value (i.e., the proposition can be shown to be true or false; Kintsch, 1994; McNamara & Magliano, 2009). Propositions consist of predicate (argument, argument), i.e., $p(x, y)$, where the arguments fill slots determined by the predicate. As an example, the sentence *He hands the book to the student* would comprise a predicate (hand) and three arguments including an agent (he), theme (book), and recipient (student): *hand (he, book, student)* (McNamara & Magliano, 2009).

There is substantial evidence that readers derive meaningful idea units (i.e., propositions) when reading, which supports the notion that propositions are strongly related to text comprehension. For instance, multiple propositions can strain working memory, lowering text comprehension (Kintsch & Keenan, 1973). There is also evidence that texts that have more propositions take longer to read and lead to lower text recall (Bisanz, Das, Varnhagen, & Henderson, 1992; Graesser, Hoffman, & Clark, 1980; Kintsch & Keenan, 1973). In general, models of comprehension using propositional representations are preferred over simple word-based representations. Such preferences are based on the notion that comprehension involves deriving larger units of meaning explicitly from the text (i.e., represented in terms of text-based propositions) and inferences generated by the reader that go well beyond the explicit words in the text. The coherence of a reader's representation of a text is driven by the connections established between the text-based propositions and the reader's extra-textual elaborations (McNamara & Magliano, 2009).

Methods

A number of studies indicate that text simplification can enhance text comprehension when compared to authentic texts. However, many previous studies did not examine linguistic differences in texts beyond readability formulas, and many studies did not examine language proficiency, reading proficiency, and background knowledge and their effects on text comprehension. No studies, to our knowledge, have examined comprehension using a text-retelling task. This study addresses many of these limitations by examining reading skills in 48 non-native speakers of English using a moving windows self-paced reading task followed by a text-retelling task. The effects of text type (beginning simplified texts, intermediate simplified texts, and authentic texts), language proficiency, reading proficiency, and background knowledge scores on participants' proposition recall is examined.

Participants

We collected data from 48 native speakers of Spanish enrolled at the Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) campus in San Luis Potosi, Mexico. All participants for this study studied at the high school or college level and reported ages between 15 to 24. Nineteen of the participants were female and the remaining were male ($n = 29$). All participants reported at least corrected to normal vision. Prior to data collection, all participants had taken a paper-based institutional TOEFL. The average TOEFL score for the participants was 520 ($SD = 30.741$). Descriptive statistics for the participants are provided in Table 1.

Table 1. *Descriptive statistics for 48 participants in study*

Item	Min.	Max.	Mean	SD
Age	15	24	17.708	2.153
Grade level	10	13	11.583	1.164
Grade point average (100 scale)	73	97	84.809	6.271
TOEFL scores	420	597	519.604	30.741
Background knowledge scores	7	21	13.417	3.389
Reading proficiency scores (GMRT)	7	34	20.959	7.023

Procedure

Data collection occurred in three separate sessions. In the first session, an on-line questionnaire on participant demographic information was given, followed by a background knowledge survey. The background knowledge survey assessed participants' knowledge of the topics covered within the reading passages. The second session occurred approximately one week later. In this session, the participants were administered the Gates-MacGinitie Reading Test (GMRT, MacGinitie & MacGinitie, 1989). The third session occurred on the following day. In this session, the students participated in an on-line reading experiment. This experiment assessed reading ability for both simplified and authentic texts using a self-paced, non-cumulative, moving window reading task similar to that used by Just, Carpenter, and Woolley (1982). Comprehension of these texts was assessed using true/false questions (see Crossley et al., 2014, for details of these results) and a text-retelling paradigm.

Critics of moving windows self-paced reading tasks note that the process can slow reading time

(Rayner, 1998) and does not allow readers to revisit previous sections of the text (Schotter, Tran, & Rayner, 2014). However, multiple studies have shown that moving windows self-paced reading tasks simulate eye movement data (Juola, Ward, & McNamara, 1982; Just, Carpenter, & Woolley, 1982; Rubin & Turano, 1992), although conflicting results have been reported (e.g., Kennedy & Murray, 1984; Magliano, Graesser, Eymard, Haberlandt, & Gholson, 1993).

Three text groupings were developed for the moving windows self-paced reading task. The groupings were organized so that each grouping included three authentic texts, three texts simplified to the intermediate level, and three texts simplified to the beginning level ($n = 9$). The texts in each grouping were on different topics and there was no overlap between the texts in each grouping. The texts were presented in random order. Participants were randomly, but evenly, assigned to a grouping so that each text at each level and each text was read by at least 16 participants. Thus, each participant read nine texts (three beginning level simplified texts, three intermediate level simplified texts, and three authentic texts) on nine different topics. Excerpts from texts used are presented in Appendix A.¹

Each text was presented one word at a time and the participants advanced through a text by pressing the spacebar on a computer keyboard. The words were presented sequentially and each word appeared in the same location as in a normal text. Participants were not allowed to revisit text that had already been read (i.e., participants could not reread previous text after pressing the space bar). This approach allowed for the calculation of a processing time measure (i.e., response times between spacebar presses) for each individual word (similar to word fixation rates). Prior to the actual experiment, participants were given instructions on the tasks and a practice trial.

When participants reached the end of the text, they were given time and space to type out a retelling of the text they had just read. They were then prompted to answer yes/no comprehension questions about the same text (see Crossley et al., 2014 for results from this portion of the study). The experiment was developed using E-Prime software. A font size of 14 was selected to ensure that visual factors did not affect reading speed (Legge, Pelli, Rubin, & Schleske, 1985).

Materials

Texts. The reading samples used for this study were the same as those used by Crossley et al. (2014). Briefly, the reading samples were selected from a corpus of 100 simplified news texts modified by expert material designers into three levels of simplification: advanced, intermediate, and beginning. From this corpus, we selected the initial paragraph from nine texts to use in the self-paced reading experiment. For each text, we had three versions: the authentic text, a text simplified to the intermediate level, and a text simplified to the beginning level. All texts contained the same main propositional information but not the same number of propositions (i.e., some texts elaborated on some propositions while others did not). Beginning texts contained the greatest number of propositions ($M = 20.667$, $SD = 6.082$), followed by intermediate ($M = 17.333$, $SD = 4.847$), and authentic texts ($M = 16.778$, $SD = 4.521$). However, the differences between the number of propositions at each level was not significant $F(2, 8) = 2.628$, $p > .050$.

¹ Because of copyright law, the text excerpts are truncated at under 100 words.

Only those texts that differed in linguistic features related to L2 text simplification, comprehension, and readability were selected based on an examination of the texts using the computational tool Coh-Metrix (Graesser, McNamara, Louwrese, & Cai, 2004; McNamara & Graesser, 2012). Thus, the texts differed significantly in terms of linguistic features related to meaning construction (i.e., cohesion), lexical recognition (i.e., lexical sophistication), and syntactic parsing (i.e., syntactic complexity; see Table 2). The selected indices are discussed in greater detail in Crossley et al. (2014).

Table 2. Means and statistical differences for linguistic features as a function of text level

Linguistic features	Beginning texts	Intermediate texts	Authentic texts	<i>f</i>	<i>p</i>	<i>hp2</i>
Noun overlap	0.55 (0.13)	0.27 (0.13)	0.15 (0.14)	21.371	< .001	0.640
Lexical diversity D	65.89 (14.85)	91.00 (14.93)	112.67 (24.65)	14.081	< .001	0.540
CELEX content word frequency	2.43 (0.17)	2.20 (0.20)	2.01 (0.16)	12.449	< .001	0.509
Sentence syntax similarity	0.13 (0.03)	0.10 (0.03)	0.07 (0.03)	8.707	< .001	0.420
Word familiarity	580.55 (8.99)	568.74 (10.70)	563.16 (9.63)	7.389	< .010	0.381
Word meaningfulness	368.12 (14.06)	352.70 (14.14)	346.89 (15.82)	5.989	< .010	0.333
Number of causal verbs and particles	44.12 (18.84)	32.87 (12.65)	23.70 (8.84)	4.761	< .050	0.284
LSA sentence to sentence overlap	0.23 (0.07)	0.18 (0.05)	0.14 (0.07)	4.352	< .050	0.266
Number of words	150.11 (29.80)	125.22 (21.93)	128.89 (28.25)	2.250	> .050	0.158

Notes. Standard deviation in parentheses

Background knowledge. Following the procedure described in Bellissens, Jeuniaux, Duran, and McNamara (2010), we developed a background knowledge assessment for the text topics used in this study. Thus, for each text, we developed specific multiple-choice questions that generally covered the key ideas shared among the beginning simplified, intermediate simplified, and authentic texts for each topic. The questions included the correct answer and three distracters that were thematically related (same theme but incorrect), near misses (incorrect in general), and unrelated (different theme and incorrect). For each text, we developed five text-based questions ($N = 45$). To examine item performance, we first piloted these questions with 25 undergraduate students. Based on the gathered responses, we selected 27 questions (three for each set of texts) for the final assessment. The criteria for selecting these questions were that each question did not indicate either ceiling ($M > .900$) or floor effects ($M < .250$). Descriptive statistics for the background knowledge scores for the 48 participants in this study are provided in Table 1.

Reading proficiency

All participants were administered the Gates-MacGinitie Reading Test (GMRT, Level 10/12; MacGinitie & MacGinitie, 1989). The comprehension test comprises 48 multiple-choice questions that assess students' reading comprehension ability across short passages. Each passage is associated with 2 to 6 questions. The questions assess shallow text comprehension as well as Deeper-level comprehension that require the reader to make inferences about the text. The participants were administered the standard instructions, including two practice questions, and given 35 minutes to complete the test. Descriptive statistics for the GMRT scores for the 48 participants in this study are provided in Table 1.

Retelling

After reading each text, participants were given 2.5 minutes to retell the text. Specifically, participants were provided with the following instructions: *Please retell the text you just read in the box above. You will have two and a half minutes to write. Write as much as possible and do not worry about spelling mistakes.* Retellings were typed into a textbox on the computer screen. Participants could see their retelling as they typed, but did not have access to the original text. The font for the retelling was set at 14.

Comprehension questions

After participants finished the retelling, they answered four true/false questions that corresponded to the main ideas and important details of the text (see Crossley et al., 2014, for more details about this phase of the study).

Proposition scoring

Two raters were trained to score the retellings in terms of propositions. Prior to rating, each text was divided into individual propositions, with each proposition consisting of a clause that contained a predicate and associated arguments. If a sentence comprised two clauses (i.e., *The stress of political life led him to seek comfort in food*), each clause was considered to be a proposition (i.e., *The stress of political life led him somewhere* and *He took comfort in food*). For each proposition, participants were allotted 1 point if they recalled all the main elements of the proposition and .5 point if the participants recalled some of the elements of the clause. They were given 0 points if they recalled no elements of the proposition.

Raters also coded for information provided by participants that was not explicitly located in the text (i.e., extra-textual propositions). These codes included *summaries* of the texts, text-based *inferences*, *relevant elaborations*, and *irrelevant elaborations*. Summaries were overviews of the entire text (e.g., *it talks about Argentinas dirt war* for a text about Argentina's dirty war) and are similar to paraphrases (McNamara, Levinstein, & Boonthum, 2004). Inferences were logical conclusions based on content of the texts that was not stated explicitly in the text (e.g., *the PepsiCo workers think that they are improving very well* for a text that stated Pepsi overtook Coke in sales but did not explicitly state that Pepsi was improving). Text inferencing allows readers to form more cohesive representations of texts that are global in nature (Kintsch, 1998). Inferencing is also more likely to occur for better comprehenders (Oakhill, 1984). Relevant elaborations occurred when the reader went beyond the text, but the idea was still related to the text topic (e.g., *And that men should not help women*) for a text that discusses changes in attitudes in Spain about household responsibilities, but does not include opinions on the topic. The production of relevant elaborations is associated with improved learning and comprehension (Bransford & Johnson, 1972; Pressley et al., 1992; Spilich, Vesonder, Chiesi, & Voss, 1979). The code *irrelevant elaborations* was used to classify ideas produced by the participants that were off-topic (e.g., *trying to disipate those interrogants* for a text about the northern lights).

In total, the two raters scored 1009 propositions produced by the 48 participants. Overall, the raters agreed on the classification for 992 of the propositions (inter-rater reliability = 98.3%). For

those instances where raters did not agree (the remaining 17 propositions), the raters adjudicated differences until agreement was reached.

Statistical Analyses

We used R (R Core Team, 2013) and *lmer* (Kuznetsova, Brockhoff, & Christensen, 2014) to perform linear, “mixed effects analyses” of the relationship between text level and proposition recalls and text level and extra-textual propositions. For each analysis, we developed two models using *lmer*. The first model was a baseline model that predicted propositional recall (or extra-textual propositions) including TOEFL, background knowledge, and GMRT scores as fixed effects and subjects as a random effect. The second model was a full model that was similar to the baseline model but included the fixed effect of text level (beginning, intermediate, authentic texts). To compare the two models, we used a log likelihood ratio test to obtain p-values for the full model for the effect in question (i.e., text level) against the baseline model without the effect in question. For all models we report the coefficients of the predictors, their standard error, and derived p-values from the t-values for each of the factors in the model.

Results

Assumptions

Visual inspections of residual plots did not reveal any obvious deviations from homoscedasticity or normality. Correlations among the fixed effects reported no strong multicollinearity (defined as $r > .70$).

Propositions Recalled

Baseline model. The linear mixed-effects model revealed a statistically significant main effect for reading proficiency, $t(43.990) = 3.084, p < .010$. The coefficients indicated that an increase in GMRT score of 1 would lead to a gain of .006 propositions recalled (or about a 1% increase). No other fixed effects demonstrated significant results. See Table 3 for the coefficients, standard errors, and p values for each fixed effect in the model.

Table 3. *Baseline linear mixed effects model for number of text-based propositions recalled*

Fixed effect	Coefficient	Standard error	t	p
Reading proficiency (GMRT)	0.006	0.002	3.084	0.004
Background knowledge	0.003	0.003	1.018	0.314
Language proficiency (TOEFL)	0.000	0.000	0.395	0.695

Note. Coefficients indicate change in text-based propositions recalled

Full model. The linear mixed-effects model revealed a statistically significant main effect for text level, $t(94.860) = -5.580, p < .001$; and for reading proficiency, $t(44.000) = 3.084, p < .010$. The coefficients indicated that moving from a lower text level (e.g., a beginning level simplified text) to a higher text level (e.g., an intermediate level simplified text) would result in a gain of -.038

propositions recalled or 4% fewer propositions. Like the baseline analysis, an increase in GMRT score of 1 would lead to a gain of .006 propositions recalled. No other fixed effects demonstrated significant results. See Table 4 for the coefficients, standard errors, and p values for each fixed effect in the model.

Table 4. *Full linear mixed effects model for number of text-based propositions recalled*

Fixed effect	Coefficient	Standard error	t	p
Text level	-0.038	0.007	-5.580	0.000
Reading proficiency (GMRT)	0.006	0.002	3.084	0.004
Background knowledge	0.003	0.003	1.018	0.314
Language proficiency (TOEFL)	0.000	0.000	0.395	0.695

Note. Coefficients indicate change in text-based propositions recalled

Comparison between models. There was a significant difference between the two models indicating that text type significantly affected the number of propositions recalled, $\chi^2(1)=27.217$, $p < .0001$, beyond reading proficiency alone. Descriptive statistics for propositions recalled as a function of text level are provided in Table 5.

Table 5. *Mean proportion of text-based propositions recalled as a function of text level*

Text level	Propositions recalled
Beginning	0.2447 (0.108)
Intermediate	0.2181 (0.101)
Authentic	0.1696 (0.098)

Note. Standard deviation in parentheses

Extra-Textual Propositions Produced

Baseline model. The linear mixed-effects model revealed a statistically significant main effect for language proficiency, $t(44) = -2.510$, $p < .050$. The coefficients indicated that an increase in TOEFL score of 1 would lead to a gain of .012 extra-textual propositions produced (or a half a percent increase). No other fixed effects demonstrated significant results. See Table 6 for the coefficients, standard errors, and p values for each fixed effect in the model.

Table 6. *Baseline linear mixed effects model for number of extra-textual propositions produced*

Fixed effect	Coefficient	Standard error	t	p
Language proficiency (TOEFL)	-0.012	0.005	-2.510	0.016
Background knowledge	-0.027	0.035	-0.762	0.450
Reading proficiency (GMRT)	0.013	0.020	0.642	0.524

Note. Coefficients indicate change in extra-textual propositions recalled

Full model. The linear mixed-effects model revealed a statistically significant main effect for text level, $t(94.990) = -3.410$, $p < .001$; and for language proficiency, $t(44.000) = 2.510$, $p < .050$. The coefficients indicate that moving from a lower text level (e.g., a beginning level simplified

text) to a higher text level (e.g., an intermediate level simplified text) would result in the production of an additional .252 extra-textual propositions (or a 25% increase in the number of propositions recalled). Like the baseline analysis, an increase in TOEFL score of 1 would lead to the production of an extra .012 extra-textual propositions. No other fixed effects demonstrated significant results. See Table 7 for the coefficients, standard errors, and p values for each fixed effect in the model.

Table 7. *Full linear mixed effects model for number of extra-textual propositions produced*

Fixed effect	Coefficient	Standard error	t	p
Text level	0.252	0.074	3.410	0.000
Language proficiency (TOEFL)	-0.012	0.005	-2.510	0.016
Background knowledge	-0.027	0.035	-0.762	0.450
Reading proficiency (GMRT)	0.013	0.020	0.642	0.524

Note. Coefficients indicate change in extra-textual propositions recalled

Comparison between models. There was a significant difference between the two models indicating that text type significantly affected the number of extra-textual propositions, $\chi^2(1) = 11.083$, $p < .0001$, beyond language proficiency alone. Descriptive statistics for extra-textual propositions produced as a function of text level are found in Table 8.

Table 8. *Mean number of extra-textual propositions produced as a function of text level*

Text level	Extra-textual propositions
Beginning	1.323 (0.992)
Intermediate	1.583 (0.983)
Authentic	1.826 (1.117)

Note. Standard deviation in parentheses

Post-hoc Analysis: Individual Additional Proposition Scores

We conducted post-hoc analyses to investigate differences in the types of extra-textual propositions produced (summary, inference, relevant elaborations, and irrelevant elaborations) among text levels. The data for the individual types of proposition scores were not normally distributed (see Table 9 for means and standard deviations for each group). We thus conducted a Friedman's two-way analysis of variance by ranks followed by related-samples Wilcoxon signed ranked tests to assess differences as a function of text level². For the summary and inference extra-textual propositions, there were no significant differences as a function of text level [summary: $\chi^2(2) = 5.021$, $p > .050$; inference: $\chi^2(2) = 1.860$, $p > .050$]. For the relevant elaborations, there were significant differences as a function of text level, $\chi^2(2) = 14.199$, $p < .001$, and follow-up analyses demonstrated significant differences between beginning and intermediate texts ($Z = 3.080$, $p < 0.10$) and beginning and authentic texts in the number of relevant elaborations made ($Z = -3.768$, $p < 0.001$) with fewer relevant elaborations in beginning level texts as compared to intermediate and authentic texts. No differences in the number of relevant elaborations were reported between intermediate and authentic texts. For the irrelevant

² A Bonferroni correction was made to adjust for multiple comparisons ($\alpha = .0166$)

elaborations, there was not a significant effect of text level, $\chi^2(2) = 3.219, p > .050$, and there were no significant differences between beginning and intermediate text; but there was a marginal difference between beginning and authentic texts ($Z = 2.787, p < 0.10$), and a significant difference between intermediate and authentic texts in the number of irrelevant elaborations made ($Z = -3.023, p < 0.001$), with greater incidences of irrelevant elaborations produced in authentic texts.

Table 9. Mean number of extra-textual propositions produced as a function of text level and type of extra-textual proposition

Type	Beginning texts	Intermediate texts	Authentic texts
Summary	0.493 (0.575)	0.451 (0.544)	0.597 (0.670)
Inferences	0.462 (0.567)	0.458 (0.399)	0.458 (0.410)
Relevant elaboration	0.257 (0.390)	0.514 (0.591)	0.639 (0.618)
Irrelevant elaboration	0.132 (0.225)	0.181 (0.257)	0.438 (0.649)

Note. Standard deviation in parentheses

Discussion

This study examined the effects of text simplification on L2 readers' text recall. To this end, a text-retelling procedure was used following a moving windows self-paced reading task, including authentic texts and simplified to beginning and intermediate levels. The use of a retelling approach afforded the opportunity to examine the production of text-based propositions, as well as extra-textual elements such as text summarization, inferences, relevant elaborations, and irrelevant elaborations. Unlike previous studies that have examined the effects of text simplification on comprehension, this study also examined the influence of language and reading proficiency and background knowledge on L2 readers' text recall.

Overall, the results of this study show that reading proficiency and text level lead to a greater number of propositions recalled in the text-retelling task. L2 readers with higher reading proficiency scores (based on the GMRT) recalled more text-based propositions. The results also indicated that beginning level texts lead to a greater number of propositions recalled. Notably, a full linear effects model, including both reading proficiency and text level, performed significantly better than a model with reading proficiency alone, indicating that text level was a stronger predictor of propositional recall than reading proficiency.

The models also reported differences for the number of extra-textual propositions produced by the readers based on language proficiency and text level. These results indicate that lower proficiency L2 learners (as assessed by TOEFL scores) produce a greater number of extra-textual propositions. In addition, readers of authentic texts produce a greater number of extra-textual propositions. A full model, including both language proficiency and text level performed significantly better than a model with language proficiency alone, indicating that text level was a stronger predictor of the production of extra-textual propositions.

Post-hoc analyses of the extra-textual propositions included within the retellings revealed no differences in the production of summary propositions, inferences, and irrelevant elaborations as

a function of text level. However, there were differences in the production of relevant elaborations with increasing text levels leading to a greater number of relevant elaborations. Specifically, intermediate and authentic texts led to more relevant elaborations than beginning texts. This is an important consideration because research indicates that relevant elaborations can be linked with improved learning and text comprehension (Bransford & Johnson, 1972; Pressley et al., 1992; Spilich et al., 1979). Specifically, relevant elaborations may indicate that information in the text is linked to information already known by the reader. Thus, as a reader makes connections between the text and prior knowledge, a more coherent and stable mental representation of the text may emerge (Kintsch, 1998).

From a linguistic perspective, the results of this study demonstrate that simplified texts led to greater propositional recall. Therefore, texts that had greater cohesion (more semantic similarity, noun overlap, word repetition, syntactic similarity, and causality) and less lexical sophistication (more frequent words, more familiar words, more meaningful words) led to greater propositional recall, but a lower number of relevant elaborations. Conversely, those texts that were less cohesive and had greater lexical sophistication led to a greater number of relevant elaborations. These findings cautiously support the modification of texts using both structural and intuitive approaches in that these modifications led to greater recall of propositions (and likely better comprehension). However, these modifications appear to lead to a decreased number of relevant elaborations, which are linked with improved learning and text comprehension.

The results also have important implications for individual differences in L2 readers and how these differences can influence propositional recall and the production of extra-textual propositions. For instance, reading ability is a significant predictor of propositional recall and should be factored into pedagogical considerations (e.g., in the selection of reading texts and reading comprehension assignments). While text simplification affects text recall, it is likely more beneficial for low proficiency readers as compared to high proficiency readers. Likewise, language proficiency is a significant predictor of the production of extra-textual information. Extra-textual relevant elaborations have been linked to improved learning and text comprehension and the greater production of these elaborations by lower proficiency L2 learners indicates that low proficiency learners may develop more connections between the text and their prior knowledge in order to create greater textual coherence. Such connections should likely be encouraged for lower level L2 learners.

Conclusion

This study demonstrates that text simplification does lead to greater propositional recall. In addition, text simplification leads to fewer relevant elaborations on the part of readers, which may indicate that beginning level texts do not require, to the same degree, links between prior knowledge and the propositions in the text. In general, the study shows benefits for simplified texts but also contains caveats for their use in the L2 classroom (i.e., simplified texts lead to greater comprehension gains, but readers appear to generate fewer inferences).

The results also indicate the development of additional research questions. For instance, knowing that simplified texts may lead to fewer relevant elaborations and knowing that relevant

elaborations are related to stronger mental representations of texts, the need for delayed tests of text comprehension become apparent. Authentic texts may lead to similar or greater gains in propositional recall after a time delay (i.e., not after an immediate test of comprehension), particularly if the authentic texts are inducing increased active processing on the part of the reader (McNamara & Kintsch, 1996). The results also call for additional research into how reading proficiency and language proficiency interact with comprehension in L2 readers. These individual differences are often not controlled for in behavioral studies that investigate text simplification, but they have a strong influence on text comprehension, as well as, the production of relevant elaborations for L2 readers. Another individual difference not co-varied in this study that may provide an additional research route is working memory.

We recognize that the L2 language instructor and researcher may be left in a quandary in terms of a take-away message from the results of this study. Indeed, the results are somewhat complex. Nonetheless, in light of this study and the bulk of evidence from prior research, instructors and researchers can be assured that simplified texts improve recall for L2 readers. However, if the pedagogical goal emphasizes *learning* from the text rather than simple text recall, then the less cohesive and more lexically challenging authentic texts induce more extra-textual inferences, which are associated with enhanced learning from text (McNamara et al., 1996; McNamara & Kintsch, 1996).

Acknowledgments

This research was supported in part by the Institute for Education Sciences (IES R305A080589 and IES R305G20018-02). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES.

References

- Allen, D. (2009). A study of the role of relative clauses in the simplification of news texts for learners of English, *System*, 37, 585–599. doi:10.1016/j.system.2009.09.004
- Bellissens, C., Jeuniaux, P., Duran, N. D., & McNamara, D. S. (2010). A text relatedness and dependency computational model: Using Latent Semantic Analysis and Coh-Metrix to predict self-explanation quality. *Studia Informatica Universalis*, 8, 85–125.
- Bisanz, G. L., Das, J. P., Varnhagen, C. K., & Henderson, H. K. (1992). Structure and components of reading times and recall for sentences in narratives: Exploring changes with age and reading ability. *Journal of Educational Psychology*, 84, 102–114. doi: 10.1037/0022-0663.84.1.103
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, 11, 717–726.
- Buswell, G. T. (1922). *Fundamental reading habits: A study of their development*. Chicago: University of Chicago Press.
- Carrell, P. L. (1987) Readability in ESL. *Reading in a Foreign Language*, 4, 21–40.

- Crossley, S. A., Allen, D., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language, 23*, 84–102.
- Crossley, S. A., Allen, D., & McNamara, D. S. (2012). Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research, 16*, 89–108.
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly, 42*, 475–493. DOI: 10.1002/j.1545-7249.2008.tb00142.x
- Crossley, S. A., Louwse, M. M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal, 91*, 15–30. doi: 10.1111/j.1540-4781.2007.00507.x
- Crossley, S. A., & McNamara, D. S. (2008). Assessing second language reading texts at the intermediate level: An approximate replication of Crossley, Louwse, McCarthy, and McNamara (2007). *Language Teaching, 41*, 409–229. doi: 10.1017/S0261444808005077
- Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language, 26*, 92–113.
- Davison, A., & Kantor, R. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly, 17*, 187–209. doi: 10.2307/747483
- Day, R. R., & Park, J. (2005). Developing reading comprehension questions. *Reading in a Foreign Language, 17*, 60–73.
- Gates-MacGinitie Reading Tests. (1989). *Technical Report for Gates-Macginitie Reading Tests forms K and L*. Chicago, Illinois: Riverside Publishing
- Graesser, A. C., Hoffman, N. L., & Clark, L. F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior, 19*, 135–151.
- Graesser, A., McNamara, D., Louwse, M., & Cai, Z. (2004). *Coh-Matrix: Analysis of text on cohesion and language*. *Behavioral Research Methods, Instruments, and Computers, 36*, 193–202. doi: 10.3758/BF03195564
- Hill, D. (1997). Survey review: Graded readers. *ELT Journal, 51*, 57–81.
- Juola, J. E., Ward, N. J., & McNamara, T. (1982). Visual search and reading of rapid serial visual presentations of letter strings, words, and text. *Journal of Experimental Psychology: General, 111*, 208–227. doi: 10.1093/elt/51.1.57
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.
- Kennedy, A., & Murray, W. S. (1984). Inspection times for words in syntactically ambiguous sentences under three presentation conditions. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 833–849. doi:10.1037//0096-1523.10.6.833
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95*, 163–182. doi: 10.1037/0033-295X.95.2.163
- Kintsch, W. (1994). The psychology of discourse processing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 721–740). San Diego, CA: Academic Press.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge; New York, NY, USA: Cambridge University Press.
- 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, 257–274. doi:10.1016/0010-0285(73)90036-4
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. R package version 2.0-11
- Legge, G. E. (2006) *Psychophysics of reading in normal and low vision*. Mahwah, NJ: Lawrence Erlbaum.
- Little, D., Devitt, S., & Singleton, D. (1989). *Learning foreign languages from authentic texts: Theory and practice*. Dublin, Ireland: Authentik.
- Long, M., & Ross, S. (1993). Modifications that preserve language and content. In M. L. Tickoo (Ed.), *Simplification: Theory and application* (pp. 29–52). Singapore: SEAMEO Regional Language Center.
- MacGinitie, W., & MacGinitie, R. (1989). *Gates-MacGinitie Reading Test* (3rd Edition). Riverside Publishing.
- Magliano, J. P., Graesser, A. C., Eymard, L. A., Haberlandt, K., & Gholson, B. (1993). Locus of interpretive and inference processes during text comprehension: A comparison of gaze durations and word reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 704–709. doi:10.1037//0278-7393.19.3.704
- McNamara, D., & Graesser, A. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 188–205). Hershey, PA: IGI Global.
- McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.
- McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, 22, 247–288.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1–43. doi: 10.1207/s1532690xcil401_1
- McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers*, 36, 222–233. doi: 10.3758/BF03195567
- McNamara, D. S., & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation*. New York, NY: Elsevier Science.
- Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, 54, 31–39. doi: 10.1111/j.2044-8279.1984.tb00842.x
- Oh, S. (2001). Two types of input modification and EFL reading comprehension: Simplification versus elaboration. *TESOL Quarterly*, 35, 69–96. doi: 10.2307/3587860
- Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist*, 27, 91–109. doi: 10.1207/s15326985ep2701_7
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. doi:10.1037/0033-2909.124.3.372
- Rubin, G. S., & Turano, K. (1992). Reading without saccadic eye movements. *Vision Research*, *32*, 895–902.
- Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science*, *25*, 1218-1226. DOI: 10.1177/0956797614531148
- Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, *41*, 159–189. doi: 10.3102/00028312041001159
- Simensen, A. M. (1987). Adapted readers: How are they adapted? *Reading in a Foreign Language*, *4*, 41–57.
- Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 275–290. doi: 10.1016/S0022-5371(79)90155-5
- Tweissi, A. I. (1998). The effects of the amount and the type of simplification on foreign language reading comprehension. *Reading in a Foreign Language*, *11*, 191–206.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Yano, Y., Long, M., & Ross, S. (1994). Effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, *44*, 189–219. doi: 10.1111/j.1467-1770.1994.tb01100.x

Appendix A

Example beginning level text excerpt

Pepsi-Cola and Coca-Cola are probably the most famous soft drinks in the world. For years Coca-Cola has been number one. Sales of Coca-Cola have always been much higher than sales of Pepsi-Cola. However, on December 12, 2005 something changed. For the first time ever, the Pepsi-Cola company was worth more on the stock market than Coca-Cola. Pepsi-Cola's market value was \$98.4bn on December 12. Coca-Cola's market value was \$97.9bn. Coca-Cola was suddenly number two, not number one.

Example intermediate level text excerpt

On December 12 people at Pepsi Cola's headquarters were probably drinking champagne rather than cola. By the end of trading on Wall Street that day, the company's market value reached \$98.4bn while the market valued Pepsi Cola's rival Coca-Cola at \$97.9bn. For the first time in the history of the two companies, PepsiCo was valued more highly than its old arch enemy. It was mainly a symbolic event but it was a powerful symbol - and one that remained over the days that followed. The "real thing" is suddenly second-best.

Example authentic text excerpt

The fizzy drink of choice at PepsiCo on December 12 was more likely to have been champagne than cola.

By the end of trading on Wall Street that day, the company's market capitalization reached \$98.4bn - and the market valued rival Coca-Cola at \$97.9bn. For the first time in the history of the two companies, PepsiCo was valued more highly than its old arch enemy. It was chiefly a symbolic shift, but what a symbol - and one that has persisted over ensuing days. The "real thing" is suddenly second best.

About the Authors

Scott Crossley is an Associate Professor at Georgia State University. His interests include the application and development of natural language processing tools in educational technology. He has published articles on the use of natural language processing tools to examine lexical acquisition, writing proficiency, reading comprehension, discourse processing, language assessment, and automatic feedback in intelligent tutoring systems. E-mail: sacrossley@gmail.com

Danielle McNamara is a Professor at Arizona State University. Her work involves the theoretical study of cognitive processes as well as the application of cognitive principles to educational practice. Her current research ranges a variety of topics including text comprehension, writing strategies, building tutoring technologies, and developing natural language algorithms. E-mail: dsmcnamara1@gmail.com