This is an Accepted Manuscript of an article published by Taylor & Francis in Reading Psychology on 23 Feb 2021 (Published online), available online: http://www.tandfonline.com/10.1080/02702711.2021.1912867.

Predicting expository text processing: causal content density as a critical expository text metric

D. Jake Follmer, Ping Li, & Roy B. Clariana

Abstract

In this investigation, we examine the contribution of intrinsic content density (ICD) to measures of expository text processing. In Studies 1 and 2, the factor structure of select text density metrics was examined and refined using two text samples (*N*s=150) randomly selected from an expository text corpus. Scores on the ICD measure based on the entire text sample (*N*=300) explained unique variance in readability and text easability. In Study 3, ICD predicted adults' text ratings of interest and ease of comprehension above and beyond established easability measures. Participants' text familiarity moderated the relation between ICD and ease of comprehension, revealing a density-facilitative effect for participants more familiar with the text content. Finally, in Study 4, measures of text difficulty, processing, and comprehension were obtained from adult readers using 10 researcher-constructed science texts; evidence of descriptive density effects on each measure was obtained. Implications for future research are discussed.

Keywords: expository text; text processing; density; readability; comprehension; textual analysis

Cite as: Follmer, D. J., Li, P., & Clariana, R. B. (2021). Predicting expository text processing: causal content density as a critical expository text metric. *Reading Psychology, 42*(6), 625-662. https://www.tandfonline.com/doi/full/10.1080/02702711.2021.1912867

Predicting expository text processing:

Causal content density as a critical expository text metric

Models of and approaches to understanding text processing have benefited from a focus on features of text that are believed to support readers' comprehension processes. Such a focus on text features and characteristics, including, as examples, readability, concreteness, cohesion, and syntactic simplicity, has been substantially augmented by advancements in natural language processing and computer-based analyses of text (e.g., Crossley, Allen, & McNamara, 2011; Graesser & McNamara, 2011). The use of these text metrics, representing quantitative indices of unique but overlapping characteristics of text, has been applied to text selection, scaling, and matching with the aim of supporting readers' skills and strategy use as they engage with complex and varied text (e.g., Crossley, Skalicky, Dascalu, McNamara, & Kyle, 2017). Much of this focus on text characteristics has historically centered on broad measures of readability (such as instantiations of Flesch reading ease and Lexile levels of text), which have been used for decades by educators, psychologists, and practitioners to inform the selection of texts for readers across the learning lifespan (Graesser, McNamara, & Kulikowich, 2011).

In contrast to the focus on readability and related measures, recent work has demonstrated the value in and utility of incorporating more nuanced indices of linguistic features of text, extending beyond traditional metrics that emphasize either word and sentence length or cloze task approaches to appraising text difficulty. One promising feature of text is conceptualized as density (see, for example, similar work on thickness by Alterman & Bookman, 1990). Density has been used to refer to the incidence of specific syntactic patterns and phrases in text that are presumed to affect text processing; for example, in supporting the degree of inferred relations that are required to be made between terms and concepts in text (see similar

work by Vidal-Abarca, Martínez, & Gilabert, 2000). Several disparate approaches for measuring text density are available via natural language processing and related linguistic tools (e.g., Graesser et al., 2011). Yet, a component-based approach to the assessment of text density – one that aggregates measures of the density of text features in expository text – is lacking in contemporary analyses of text characteristics (see analogous work on text easability components; Graesser & McNamara, 2011).

Early conceptions of text density – often instantiated as syntactic density – emphasized the tabulation of linguistic structures in text and clause length (e.g., Belanger, 1978; Golub & Kidder, 1974). More modern approaches to assessing density include specific part-of-speech incidence metrics, such as those representing the incidences of noun and verb phrases (see also similar research examining semantic and contextual diversity; Hills, Maouene, Riordan, & Smith, 2010; Hsiao & Nation, 2018). Despite advances in methods for computing the density of word classes, the relations among different density metrics are in many ways not well understood. Further, contemporary linguistic accounts often position syntactic pattern density as a feature of text that is indicative of text complexity. In other words, an overarching belief is that text with higher density (e.g., of noun and verb phrase information) is likely to be more informationally compact and, thus, harder to process. Yet, recent work (Authors, 2020) has illuminated facilitative effects of text density on readers' comprehension judgments during reading and study tasks, suggesting that features of syntactic pattern density may be supportive of readers' processing of text and their corresponding judgments of text understanding. Based on previous research (McCarthy & Jarvis, 2007), this work aims to present a comprehensive empirical examination of the effect of text syntactic pattern density on measures of expository text processing.

The expository text genre is a form of communication that intends to provide factual information to a defined audience for a defined purpose (Wang & Cho, 2010). Expository texts typically follow one of five formats: cause and effect, compare and contrast, descriptions, problem and solution, and sequences (Meyer & Poon, 2001). In English expository text, this information is carried as structured sequences of propositions (sentences), comprised of a subject (noun), verb, and object (noun), S-V-O. Thus, measures of the density of noun and verb syntactic patterns, word types, and word phrases are likely to be relatively more fundamental components of expository text compared to other genres. Further, word forms that increase causal cohesion across propositions in text will likely reduce the complexity of the text being read (McNamara et al., 2014). An intrinsic text content density metric should, therefore, center on measures of noun and verb phrases of propositions and of cohesion across propositions.

Despite advances in computer-based analyses of text, the role of content density in promoting and constraining readers' processing of expository text remains unclear (cf. Alterman & Bookman, 1990; Pitler & Nenkova, 2008; Ravid & Berman, 2006). Grounded in existing theoretical and empirical approaches to evaluating characteristics of text (see, for example, work by Kintsch & van Dijk, 1978; Miller & Kintsch, 1980; Vidal-Abarca, Martínez, & Gilabert, 2000), in this work, we test and evaluate a measure of the density of causal text elements in expository text. We refer to this measure as intrinsic content density (ICD), and conceptualize ICD as the extent to which clauses and sentences in expository text contain causal ideas, actions, and events in expository text. We aimed to examine the utility of this measure of ICD by analyzing its ability to predict both indicators and actual measures of expository text processing across a range of texts and among adult readers (see also Crossley et al., 2017).

Models of text comprehension vary in their emphasis on the features of discourse that facilitate text processing and understanding. The construction-integration model of comprehension, for example, posits key roles of discourse features such as cohesion cues on readers' ability to derive meaning across local and global discourse elements (e.g., Kintsch & van Dijk, 1978; Kintsch, 1988, 1998). Similarly, other models of text comprehension, including the structure building (Gernsbacher, 1990, 1997) and constructionist (Graesser et al., 1994) models, acknowledge the importance of discourse features such as morphology and syntax, semantics (for coherence), referential cohesion, and situational cohesion in promoting a range of comprehension processes (McNamara & Magliano, 2009). These features of text are presumed to be instrumental in readers' construction of an accurate and organized mental representation of text, in part through their facilitation of processes such as mapping and text-based inferencing.

The construction-integration model of text comprehension (Kintsch, 1998) advances an important role of levels of representation as a mechanism for signifying textual features that facilitate text processing. In particular, the textbase is premised on propositional structure and connectedness such that ideas – indicated by predicate and argument – accurately represent the meaning of text-based information (McNamara & Magliano, 2009). The textbase is believed to overlap with and be facilitative of a reader's developing situation model to the degree that it is comprised of well-constructed propositions that are connected and organized (e.g., Zwaan & Radvansky, 1998; see, for example, Graesser & Clark, 1995 and Kintsch, 1988 for accounts of the relation between the situation model and the textbase). As a result of this emphasis on connectedness, conventionally, the construction-integration model has largely emphasized cohesion (e.g., via argument overlap) across propositions in textbase levels of representation (Kintsch, 1995). Yet, propositions – as well as text constituents more globally – can vary

considerably in complexity and dimension (e.g., Zwaan, Magliano, & Graesser, 1995). Further, as a result of the syntactic pattern density of text (i.e., specific syntactic patterns and phrases) or syntactic complexity (e.g., embedding), propositional networks may logically contain subpropositions at varied points in the textbase, representing embedded ideas that are subordinate to but facilitative of so-called core propositions (see McNamara & Magliano, 2009 for discussion of the propositional textbase). Such a structure is common to expository text where complex propositions vary based on density.

While these complex propositions are presumed to be central in influencing a reader's situation model – consisting of organized, inferred connections between and among separate parts of text that are linked with readers' domain knowledge (e.g., Graesser & Bertus, 1998; Zwaan, Magliano, & Graesser, 1995) – an approach to capturing the presence and degree of these propositions in text is largely lacking from existing work. We leverage the construction-integration model of text comprehension (Kintsch, 1988, 1998; McNamara & Kintsch, 1996) to support examination of the intrinsic content density of propositions in expository text and the role density may serve in predicting indicators and actual measures of expository text processing, including estimates of text readability and easability, ratings of text interest and comprehensibility, and measures of text comprehension.

The Current Investigation

Expository texts serve as a primary means for supporting knowledge acquisition and revision (see, for example, a review of the role of science texts in science education and learning; van den Broek, 2010). Indeed, the ability to successfully read and comprehend expository text allows individuals to modify and improve their developing mental representations of ideas, concepts, and information. Previous research has documented the importance of understanding

causality in expository text and of generating causal inferences during expository text reading (e.g., Best et al., 2005; Graesser, & Bertus, 1998). Drawing from existing work examining indicators of syntactic pattern density and situation model construction (e.g., Graesser, McNamara, Louwerse, & Cai, 2004; see, for example, causal content indices as text-based indices of situation model understanding) and using computational linguistic analysis (see Coh-Metrix; Graesser, McNamara, & Kulikowich, 2011), we examined the incidence (i.e., classified number of units) of specific text constituents, including: a) the degree of causal verbs, particles, and connectives in text, b) verb phrase information, and c) noun phrase information. Theoretically, these metrics aligned with multilevel frameworks of text comprehension that posit a key role of text constituents in shaping representations at different levels of discourse; for example, they relate to notions of propositional representation that inform the development and maintenance of the textbase and, more distally, situation model levels of text representation (Kintsch, 1998).

Collectively, these text metrics were selected to represent the degree of causal ideas and events in expository text. Our selection and analysis of text metrics was driven by several assumptions: 1) expository texts, while inherently causal in nature, vary in the degree of causal text elements, including causal verbs, particles, and connectives; 2) description, causation, and sequencing in expository text are signaled by causal verb and verb phrase information that conveys action and process; 3) the degree of text content (i.e., ideas and events) in expository text can be represented by quantifying incidences of causal text elements, including causal verbs and particles and verb phrase information; 4) the incidence of verb-based information in expository text informs, is bound to, and is constrained by the presence of noun phrase information; and 5) variability in the degree of causal text elements may exert differential effects

on indicators and actual measures of text processing, particularly when applied to expository text.

In this research, we test the utility of this content density measure by drawing on both an existing large corpus of expository texts as well as researcher-constructed expository science texts. We then conducted four studies to explore, refine, confirm, and test the contributions of the developed content density measure to indicators and actual measures of expository text processing. To evaluate the relations among content density and indicators of text processing, we analyzed expository texts based on traditional measures of readability, including Flesch reading ease and Flesch-Kincaid grade level estimates. These measures of readability are based on sentence length and the number of syllables per word; higher scores on the Flesch reading ease estimate are indicative of greater reading ease, whereas higher scores on the Flesh-Kincaid grade level estimate are indicative of greater reading difficulty (e.g., Flesch, 1948; Kincaid, Fishburne, Rogers, & Chissom, 1975). In addition, given existing criticisms of readability formulas (e.g., Begeny & Greene, 2014), including that traditional readability measures do not account for relationships among text elements (such as text cohesion; Britton & Gülgöz, 1991; Crossley et al., 2017), we also examined measures of text easability taken from Coh-Metrix. These easability metrics included word concreteness (the degree to which content words are concrete and evoke clear meaning), syntactic simplicity (the degree to which sentences use familiar syntactic structures), referential cohesion (the degree of overlap among content words and ideas in text), and deep cohesion (the degree of causal and intentional connectives in text), and provided a more nuanced examination of text ease that is grounded in theories of text comprehension (e.g., Graesser & McNamara, 2011; Kintsch, 1998). Conceptually, these measures of text easability overlap with but are distinct from the content density measure we examine in this work.

In Study 1, we randomly drew a sample of texts (*N*=150) from a large expository text corpus to examine the factor structure of select text density metrics (see the selection criteria below). As described, each of these metrics was proposed to be indicative of specific features of text content density. In Study 2, we randomly drew another distinct but related sample of texts (*N*=150) from the same corpus to confirm and refine the structure of the content density measure. We then used computed content density scores to predict scores based on traditional measures of readability as well as more contemporary measures of text easability (Flesch, 1948; Graesser et al., 2011). In this way, a major goal of this research was to examine the capability of a developed ICD measure to predict measures of text easability and explain unique variance in readability above and beyond that explained by text easability. We also examined whether the ICD measure demonstrated positive or negative associations with these readability and easability measures; this focus allowed us to evaluate possible facilitative and hindering effects of ICD on these measures of readability and easability.

Next, in Study 3, we examined the ability of ICD to predict adults' ratings of expository texts. Using a randomly selected text sample (from the larger text sample examined in Studies 1 and 2), we evaluated the contributions of ICD to adults' ratings of perceived interest in and ease of comprehension of the expository texts. Specifically, we randomly assigned a subset of the selected texts to participants and asked them to provide ratings of each text based on familiarity, interest in the text's content, and ease of comprehension (e.g., Schiefele, 1999; Schraw, Bruning, & Svoboda, 1995). We also collected information on participants' reading behaviors and attitudes. Thus, Study 3 provided an examination of the possible contributions of ICD to adults' ratings of natural or authentic expository texts. Using a similar analytic approach to Study 2, we aimed to examine the ability of ICD to predict key text ratings above and beyond 1) adults'

reading behaviors and attitudes, 2) adults' text processing (measured via text reading times), and 3) text easability measures (including word concreteness, syntactic simplicity, and deep cohesion). We also examined whether the contribution of ICD to participants' comprehension ratings varied based on their familiarity with the text content; this focus allowed for an evaluation of the degree to which an association between ICD and comprehension ratings was moderated by participants' prior knowledge of the text content.

Finally, Study 4 grounded our examination of the role of ICD in measures of text processing and comprehension based on constructed texts. Specifically, using a within-subjects design, we collected measures of difficulty, processing (as indexed by sentence-level reading times), and comprehension from adult readers based on 10 constructed science texts addressing a range of science, technology, engineering, and mathematics (STEM) concepts. We then examined descriptive effects of texts that were both low and high in ICD on readers' difficulty, processing, and comprehension. We therefore aimed to examine variation in the role of ICD in supporting readers' perceptions and processing of science texts.

Study 1

Method

Text sample. A total of 150 expository texts from the Wikipedia corpus (Shaoul & Westbury, 2010) were examined in Study 1. The use of this corpus is established in existing research (e.g., Bartunov, Kondrashkin, Osokin, & Vetrov, 2016; Bojanowski, Grave, Joulin, & Mikolov, 2017; Huang, Socher, Manning, & Ng, 2012). The Wikipedia corpus was selected for this work for several reasons. First, based on the aims of this investigation, the corpus provides a varied sample of expository texts emphasizing an expository structure across the texts. Second, the corpus collectively provides a description of a range of text content and topics (ranging from

summaries of battles and conflicts to descriptions of plant species to summaries of disorders of human functioning) which resulted in breadth of analysis of diverse terminology, syntactic complexity, and sentence structures. Third, the corpus relied on varied authors and sources and was likely to contain differences in veracity and credibility across texts. These conditions reflect the broad nature of expository texts – in multiple platforms and through multiple forms of media – to which readers across the learning timespan are often exposed and which readers must engage and evaluate to derive an accurate mental representation of text content (Kintsch, 1988, 1998; McNamara & Magliano, 2009).

The texts were preprocessed by Shaoul and Westbury (2010); links and irrelevant material were removed and texts less than 2,000 characters were trimmed from the corpus. The corpus was downloaded in the spring of 2019 (https://www.psych.ualberta.ca/~westburylab/downloads/westburylab.wikicorp.download.html). The text sample used in Study 1 (N=150)[1] was randomly selected from the total number of documents included in the corpus; the corpus is based on a total of approximately 2 million documents. The mean word length of the selected texts was 805.13 words (SD=743.37; interquartile range=522.50 words); texts ranged in length from 200 to 6,192 words (see Table 1 for a descriptive summary of the texts analyzed in Study 1). Similarly, the mean sentence length of the texts was 41.16 (SD=32.64). Thus, examined texts demonstrated adequate variability in length so as to support broad examination of the role of ICD in estimates of expository text processing.

[Insert Table 1 about here]

[1] A sample of 150 texts was selected because this text sample size was sufficient to ground the factor analytic approach conducted. This text sample size is also consistent with recent published work evaluating text processing, familiarity, and comprehension among adult readers (see Crossley et al., 2017).

Text metrics. Each of the 150 selected texts was individually analyzed on specific text metrics based on research examining features of text related to syntactic pattern density and believed to be supportive of readers' developing situation models (van Dijk & Kintsch, 1983; Graesser & McNamara, 2011; Zwaan & Radvansky, 1998). Specifically, four metrics were targeted in Study 1 that examined the following features of text: incidence of causal connectives (CNCCaus), incidence of causal verbs and particles (SMCAUSvp), incidence of noun phrases (DRNP), and incidence of verb phrases (DRVP). These incidence measures are based on the number of classified units of text per 1,000 words. Text analyses were conducted using Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004). Coh-Metrix is an automated computational system that analyzes texts along multiple characteristics and levels of language discourse (see Graesser et al., 2011 for an overview).

Collectively, these metrics were selected because they were presumed to reflect the degree of intrinsic density in the expository texts examined. In particular, the incidence of causal connectives (such as *because* or *so*) was presumed to signal elaboration on and linkages between ideas and concepts presented in the texts. Similarly, the incidence of causal verbs and particles (e.g., *make, cause*, *require*; *to*) were presumed to signal events and concepts that were causally linked in text, thus providing an indication of the degree of causal events and content described in the texts. Finally, incidence scores for both noun and verb phrases were included as representations of the degree to which the texts were informationally compact with respect to major features of syntax, including modifiers such as adjectives and adverbs, respectively (see, for example, Graesser et al., 2004, p. 6).[2] Because of the emphasis of these metrics on incidence

[2] While one might expect a strong degree of overlap between the metrics representing the incidence of causal verbs and participles and the incidence of verb phrases, the Pearson correlation coefficient between scores on the two metrics was weak to moderate in magnitude, r=0.27.

and signaling of the density of causally-related ideas and events across text, these measures were selected to represent an index of ICD in the expository texts examined.

To evaluate the relations among ICD and indicators of expository text processing, select criterion metrics were also examined. In particular, each of the texts was also analyzed based on the following additional text features: measures of text easability, including word concreteness, syntactic simplicity, referential cohesion, and deep cohesion and measures of readability, including Flesch reading ease and Flesch-Kincaid grade level estimates. Finally, given the important role of concreteness in expository text processing (Graesser et al., 2011), text easability scores based on word concreteness were also obtained for each of the texts; these word concreteness scores were systematically controlled across correlational analyses reported in Study 1.

Procedures and analytic approach. The selected texts were further processed to ensure the presence of hard returns to signal paragraph boundaries and, in general, the suitability of the texts for analysis using Coh-Metrix. To address the aim of Study 1, scores representing incidence of causal connectives, incidence of causal verbs and particles, incidence of noun phrases, and incidence of verb phrases were subjected to exploratory factor analysis (Mplus, Version 8; Muthén & Muthén, 2010; Muthén & Muthén, 2017). Given the presence of non-normality that is common among scores representing the text metrics examined, maximum likelihood rotation with robust standard error estimates was employed. Geomin rotated loadings were evaluated.

Results

Descriptive statistics as well as percentile information for each of the text metrics are presented in Table 2. Scores on the metrics demonstrated adequate variability across the texts.

[Insert Table 2 about here]

Results of the EFA suggested a one factor solution. Table 3 presents factor loading and communality information for the EFA results. All factor loadings exceeded ±0.30 and were statistically significant, *p*s < .05 (Muijs, 2011). The solution accounted for approximately 43.63% of the variability among the text density scores. Further, obtained communalities indicated, in general, that adequate variance in individual metric scores was accounted for by the one-factor solution.[3] However, one metric – incidence of causal connectives – exhibited a communality below the commonly-accepted threshold (i.e., 0.20; Kline, 2005; see Table 3), suggesting a need for further evaluation of the utility of this specific metric through analysis of additional texts. Scores on the incidence measure of noun phrase information were negatively correlated with incidence measures of both verb phrase information, r=-0.39, $p < .05$, and causal verbs and particles, r=-0.20, $p < .05$.

[Insert Table 3 about here]

To provide a preliminary evaluation of the relations among ICD and measures of readability and text easability, factor scores were computed and saved using the regression method in Mplus[4] and Pearson correlation coefficients were examined. Table 4 presents correlations among ICD and measures of readability and easability. ICD scores were moderately, positively, and significantly correlated with easability metrics based on syntactic simplicity and deep cohesion, *r*s > 0.40, *p*s < .05. A significant correlation was not observed between ICD scores and referential cohesion, $p > .05$. ICD scores were negatively and significantly correlated with Flesch-Kincaid grade level estimates and positively correlated with Flesch reading ease

[3] Communalities describe the amount (i.e., percent) of variance in specific indicators that is explained by the common factor. Communalities greater than 0.20 indicate that the indicator is well accounted for by the factor solution (Kline, 2005).
[4] This method is also referred to as the maximum *a posteriori* method.

estimates for the texts, *p*s < .05. Across correlations, the association between the measure of ICD and text easability measures ranged from small to moderate in magnitude (Cohen, 1988).

[Insert Table 4 about here]

Discussion

Notions of text density have largely been dispersed across features of syntactic pattern density. The current study provided tentative support for the utility of positioning ICD as a component-based metric that has the potential to relate with and predict indicators of text processing such as readability and easability. Specifically, based on the factor analytic evidence presented in Study 1, the text metrics examined – incidence of causal connectives, causal verbs and particles, noun phrases, and verb phrases – appear to provide related information about the ICD of expository text. As discussed, these metrics were selected because they were presumed to directly signal events, concepts, and actions that were causally linked in the expository texts examined. The findings of Study 1 support the selection and compositing of these text metrics as indicators of the density of content-based words, phrases, and events across expository text.

The structure of the four metrics examined provided evidence of differentiated loadings based on distinct discourse categories. That is, loadings based on incidence of causal connectives, causal verbs and particles, and verb phrases were signed positively on the factor, while the loading based on incidence of noun phrases was signed negatively. In some respects, this finding may be based on sentence-level constraints such that text with a greater extent of verb-based information, including modifiers and connectives, constrains available word length toward noun-based information. Irrespective of these sentence-level constraints, the results of the exploratory factor analysis and accompanying correlational analysis suggest that the ICD of text

is higher when it contains higher incidences of causal connectives, causal verbs and particles, and verb phrases relative to noun-based information. At the other end of the factor score range, the ICD of text is lower when it contains a higher density of noun phrase information, likely with embedded clause content, relative to the other indicators. Correspondingly, as the ICD of text decreases – in part because of decreases in verb-based metrics relative to increases in noun phrase incidence – scores on measures of easability and readability can be expected to increase. This latter interpretive scenario is one that conforms to traditional assumptions regarding the role of text density in text processing (e.g., Alterman & Bookman, 1990). That is, based on this ICD approach, texts with a greater ICD of noun phrase information may be more difficult to process. On the other hand, the results of Study 1 begin to provide a facilitative interpretation of the density of texts such that texts with greater ICD of verb-based metrics (including causal verbs and particles) may be supportive of key indicators of text easability and readability.

Understanding the role and effect of features of text in text processing is of considerable theoretical and psycholinguistic importance (e.g., Gernsbacher, 1990, 1997; Kintsch, 1998; McNamara & Magliano, 2009). Based on the findings of Study 1, ICD may serve as an informative text characteristic alongside well-studied and well-established text features such as referential and situational cohesion as well as morphology and syntax. In addition, the correlational findings reported suggest that the measure of ICD relates positively with both syntactic simplicity and deep cohesion as well as with Flesh reading ease as a more traditional measure of text readability. Together, these findings suggest that ICD, based on the incidence measures examined, may be supportive of readers' processing of syntactic structures in text (as represented by indices of the syntactic simplicity of text) and the extent of connection or overlap

across causal and logical relationships within text (as indicated by indices of the deep cohesion of text).

Study 2

The findings of Study 1 provide preliminary evidence for the utility of ICD as a characteristic of text that may contribute to expository text processing. Yet, the results reported are limited by the fact that they are based on a single sample of texts and rely on indicators (i.e., estimates) of text processing. Accordingly, in Study 2, we examined a second sample of expository texts based on the same text corpus used in Study 1. This approach provided us the opportunity to confirm the structure of the text metrics selected. In addition to gathering and examining confirmatory factor analytic evidence, we also used ICD scores, based on the entire text sample (N=300), to predict scores based on traditional measures of readability as well as more contemporary measures of text easability (Flesch, 1948; Graesser et al., 2011). A primary aim of Study 2 was to evaluate the ability of ICD to account for unique variance in these measures of text processing after controlling for word concreteness in the texts. Using hierarchical regression analysis, we also aimed to evaluate potential facilitative and hindering effects of ICD on these measures of readability and easability.

Method

Text sample. A second sample of 150 expository texts was randomly selected and analyzed in Study 2. In no instance did random selection result in the selection of duplicate texts across the Study 1 and Study 2 text samples. To parallel the examination of text metrics presented in Study 1, the Wikipedia corpus was again selected (Shaoul & Westbury, 2010). Table 5 presents a descriptive summary of the texts evaluated in Study 2.

[Insert Table 5 about here]

Text metrics. Consistent with Study 1, in Study 2, the following text metrics were computed and examined for each text: incidence of causal connectives; incidence of causal verbs and particles; incidence of noun phrases; and incidences of verb phrases. As in Study 1, each of the texts was also analyzed based on the following additional text features: measures of text easability, including syntactic simplicity, referential cohesion, and deep cohesion and measures of readability, including Flesch reading ease and Flesch-Kincaid grade level estimates. Text easability scores based on word concreteness were also obtained for each of the texts; these word concreteness scores were systematically controlled across analyses reported in Study 2.

Analysis and procedures. As in Study 1, texts were further processed to ensure suitability for analysis using Coh-Metrix. In the first part of Study 2, scores representing incidence of causal connectives, incidence of causal verbs and particles, incidence of noun phrases, and incidence of verb phrases were subjected to confirmatory factor analysis in Mplus (Version 8; Muthén & Muthén, 2010; Muthén & Muthén, 2017). This allowed for further evaluation of model fit based on the density measures selected to represent ICD. To more accurately estimate and appraise the loadings and contribution of each of the density metrics to ICD, the latent ICD factor was regressed on a measure of text length. Specifically, we regressed the latent ICD factor on the mean number of sentences per paragraph. This length metric was selected for two reasons. First, more generally, the ICD of expository texts may be expected to increase as paragraph length increases. Second, one of our aims with this work was to examine the contribution of ICD to measures of text easability, including referential and deep cohesion. Several of these easability components are based on the calculation of measures of overlap among and across adjacent sentences. Because expository texts vary greatly in the degree of overlap and cohesion among ideas within and across paragraphs, and because the estimation of

ICD can be expected to depend in part on the length and connectedness of text by paragraph, we decided to adjust estimated factor loadings based on this measure of text length.

Maximum likelihood rotation with robust standard error estimates was again employed. Model fit was evaluated using the following estimates and criteria: the root mean square error of approximation (RMSEA; value≤0.06); the Comparative Fit Index (CFI; value≥0.95); the Tucker Lewis Index (TLI; value≥0.95); and the standardized root mean square residual (SRMR; value≤0.08). The ratio of chi-square to *df* (χ^2/df) was also examined (criterion: < 3.00).

In the second part of Study 2, factor scores, again computed using the regression method in Mplus, were estimated to represent ICD for the entire text sample (*N*=300). ICD scores were then used in hierarchical regression analyses to separately predict both traditional measures of readability (based on Flesch reading ease and Flesch-Kincaid grade level estimates) and measures of text easability (based on syntactic simplicity and referential and deep cohesion)[5]. Text word concreteness easability scores, representing the degree of concrete compared with abstract content words in text, were entered into the first block of the regression. Then, in the second and final block, scores representing ICD were entered. Based on the number of regression models examined, a Bonferroni adjustment was made to protect against Type I error; all regression models were evaluated for statistical significance against α=0.01. Across models, evaluation of the regression models revealed that issues of multicollinearity were not present (all VIF values<2.00; all tolerance values>0.50).

Finally, to compare the predictive ability of ICD and select measures of text easability, we conducted a separate hierarchical regression analysis where we entered word concreteness scores in the first block and syntactic simplicity and deep cohesion scores in the second block of

[5] This procedure was used in lieu of structural equation modeling because a model with five continuous outcomes would not have been appropriately identified.

the regression. In the third and final block of the analysis, we entered ICD scores. Flesch reading ease scores were predicted. We also computed and present partial correlations for each of the text metrics. This procedure allowed us to: 1) evaluate of the degree of overlap and linear relation among the text easability and ICD metrics, and 2) evaluate the relative contributions of each metric to text readability.

Results

Descriptive statistics as well as percentile information for each of the density metrics in Study 2 are presented in the first part of Table 6. Scores on the metrics again demonstrated adequate variability across the texts.

[Insert Table 6 about here]

To examine the factor structure of the density metrics, a one-factor solution was specified in the CFA. As described above, the latent ICD factor was regressed on the mean number of sentences per paragraph as a measure of text length. The model fit the data well, $\chi^2(4)=4.08$, $p=.39$, CFI=1.00, TLI=1.00, RMSEA=0.01, SRMR=0.04. The ratio of χ^2 to *df* was 1.02. Text length significantly predicted ICD, $\beta=0.17$, $SE=0.08$, $p < .05$. However, somewhat consistent with the factor analytic results of Study 1, the standardized factor loading for incidence of causal connectives was low (0.19) and not statistically significant, $p > .05$. Likewise, the observed R^2 value for incidence of causal connectives was low ($R^2=0.04$) and not significant, $p > .05$; remaining R^2 values were significant, ps $< .05$. For these reasons, incidence of causal connectives was trimmed from the CFA model.

The revised CFA model, based on incidence of causal verbs and particles, incidence of noun phrases, and incidences of verb phrases, fit the data well, $\chi^2(1)=0.73$, $p=.39$, CFI=1.00, TLI=1.02, RMSEA=0.001, SRMR=0.01. The ratio of χ^2 to *df* was 0.73. The final CFA model is

presented in Figure 1. Factor loadings based on the density metrics were significant (>±0.40); observed R^2 values ranged from 0.24 to 0.36, *p*s < .05. Text length again predicted ICD, β=0.21, *SE*=0.08, $p < .05$, suggesting value in adjusting ICD scores based on the mean number of sentences per paragraph in the texts.

[Insert Figure 1 about here]

Next, to obtain factor scores representing ICD for the entire text sample (*N*=300), a third CFA with the latent ICD factor regressed on text length was conducted[6]. Similar to the CFA model based on the second text sample, the model based on the entire text sample fit the data well, $\chi^2(1)=0.84$, $p=.36$, CFI=1.00, TLI=1.01, RMSEA=0.001, SRMR=0.01; the ratio of χ^2 to *df* was 0.84. Descriptive statistics and percentile information for individual density metrics for the entire text sample are also presented in Table 6. The ICD factor scores demonstrated approximate normality as evidenced by skewness, 0.12, and kurtosis, -0.01, statistics as well as both the Kolmogorov-Smirnov, $D(300)=0.03$, $p=.20$, and Shapiro-Wilk, $W(300)=0.99$, $p=.84$, tests. Scores on the measure of ICD and sentence length, as indexed by the mean number of words in each sentence within text, were weakly correlated, $r=0.16$, $p < .05$, suggesting that the ICD of expository text was not primarily driven by sentence length.

To address the second aim of Study 2, factor scores representing ICD were used to predict indicators of text processing. Results of the regression analyses are presented in Table 7. Scores based on the measure of ICD predicted both syntactic simplicity and deep cohesion, *p*s < .01, accounting for unique variance in both easability metrics above and beyond text concreteness (ΔR^2=0.07 and 0.11, respectively). Somewhat consistent with the correlational findings of Study 1, ICD did not significantly predict referential cohesion, $R^2=0.01$, $p > .05$.

[6] This model likewise excluded incidence of causal connectives.

Results of the analysis indicated a predictive role of ICD in both the degree of familiar syntactic structures in text and the extent of connection or overlap across causal and logical relationships within text. Combined with the measure of text concreteness, ICD accounted for 17% and 14% of the variance in syntactic simplicity and deep cohesion, respectively (R^2=0.17 and 0.14).

[Insert Table 7 about here]

Scores based on the measure of ICD also predicted traditional measures of text readability. Specifically, density scores explained unique variance in estimates of Flesch reading ease (ΔR^2=0.09) and the Flesch-Kincaid grade level (ΔR^2=0.09), *p*s < .001. Across readability measures, findings indicated that as ICD increased, readability correspondingly improved. Combined with measures of text length and concreteness, ICD accounted for 21% and 15% of the variance in reading ease and grade level estimates, respectively (R^2=0.21 and 0.15).

[Insert Table 8 about here]

Next, hierarchical regression analysis was conducted to compare the predictive utility of measures of text easability and the measure of ICD. Summary results of the regression analyses are presented in Table 8. Evaluation of variance inflation factor (<2.00) and tolerance (>0.40) estimates indicated that issues with multicollinearity were not present in the model (all condition index values were also less than 15.00). Consistent with established research supporting use of measures of text easability (Graesser et al., 2011), the measures of syntactic simplicity and deep cohesion accounted for unique variance in text readability after accounting for word concreteness, $F(2, 296)=8.11$, $\Delta R^2=0.05$. In the third block of the regression, ICD scores significantly predicted text readability, $b=1.14$, $SE=0.25$, $t=4.48$, $p < .001$, explaining unique variance in readability after accounting for the effects of word concreteness, syntactic simplicity, and deep cohesion, $F(1, 295)=20.08$, $\Delta R^2=0.05$. The partial correlation between ICD and

readability, r=0.25, was greater than the partial correlation between syntactic simplicity and readability, r=0.10, and between deep cohesion and readability, r=0.02. The overall regression model was statistically significant, $F(4, 295)=21.03$, and explained a total of 22% of the variance in reading ease, $R^2=0.22$.

Discussion

The findings of Study 2 provide additional evidence for the predictive role of ICD in indicators of expository text processing. In particular, the results of Study 2 largely corroborated those obtained in Study 1, suggesting that the density metrics examined – incidence of causal verbs and particles, incidence of noun phrases, and incidences of verb phrases – support a one-factor structure representing the ICD of expository text. After some refinement to the final statistical model, factor scores representing ICD accounted for significant and unique variance in measures of both text readability and easability. Evaluation of a competing statistical model also revealed that ICD scores accounted for unique variance in readability after accounting for the effects of syntactic simplicity and deep cohesion.

Evaluation of the CFA in Study 2 indicated that the incidence of causal connectives – a metric representing the degree of connectives that causally link ideas and clauses in text (such as *because* or *so*; Louwerse, 2001) – did not contribute to the ICD factor. To some degree, the inclusion of causal connectives in the factor overlaps with the presence of causal verbs and particles because causal particles (such as *to*) signal how events and processes in expository text are connected, particularly in the presence of causal verbs (such as *make* and *require*). Previous research has found that causal connectives are indicative of causal cohesion and, accordingly, discriminate texts that are low and high in cohesion (e.g., Graesser et al., 2004; Louwerse, 2001). Given that the focus of our ICD measure is on incidence (as a representation of density) and not

on cohesion, it may be that the measure of causal connectives is most relevant to and better placed as a metric contributing to the causal cohesion of texts (see Table 1 of Graesser et al., 2011).

Study 3

While the results of Study 2 provide additional evidence for the facilitative effects of ICD on text processing, they rely entirely on estimates or indicators of text processing. That is, the results of and relations based on Study 2 are inferred and are not based on actual measures of learners' processing of expository text. Accordingly, in Study 3, we obtained ratings and measures of text difficulty and processing from a broad sample of adult readers using a subsample of the same texts examined in Studies 1 and 2. Specifically, in Study 3, we randomly selected 90 texts from the entire text sample analyzed in Study 2 (representing 30% of the expository text sample) and collected adults' ratings of perceived interest in (e.g., Schiefele, 1999) and ease of comprehension of (e.g., Schraw et al., 1995) the texts. We also collected a rating of adults' familiarity in the texts as well as basic reading behaviors (e.g., number of books read in a year, enjoyment of reading) for each participant.

The primary aim of Study 3 was to evaluate the ability of ICD to account for unique variance in actual measures of text processing. Accordingly, we again employed hierarchical linear regression and controlled for adults' reading behaviors and familiarity with the texts that were rated. As in Study 2, we also controlled for important easability metrics for the texts, including concreteness, syntactic simplicity, and deep cohesion; this allowed for evaluation of multicollinearity between the text metrics as well as of the relative predictive ability of each metric. Given our focus on examining the ability of ICD to explain unique variance in adults'

text ratings, we also evaluated partial correlations between each of the text metrics examined in the regression analyses.

Method

Participants. A total of 90 adults participated in Study 3. Participants were sampled from Prolific, an online, research-focused participant recruitment platform supported by previous research as providing high-quality data on varied participants (e.g. Palan & Schitter, 2018; Peer et al., 2017). The mean age of participants was 32.46 (*Mdn*=31.00; *SD*=11.48). Approximately 40.0% of participants identified as male while 58.9% identified as female; one participant (1.1%) identified as non-binary. With regard to race and ethnicity, approximately 67.8% of participants identified as White, 6.7% identified as Black, 7.8% identified as Asian, 2.2% identified as Native American, and 5.6% identified as Hispanic; approximately 8.9% of participants identified as more than one race. A total of 13.3% of participants completed a high school diploma or equivalent, 25.6% completed some college, 12.2% completed an associate's degree, 33.3% completed a bachelor's degree, and 14.4% completed a master's degree; in addition, one participant (1.1%) completed a doctoral degree.

Text sample. A total of 90 expository texts was randomly selected from the larger text sample examined in Study 2 (i.e., 30% of the text sample). The mean length of the text sample examined in Study 3 was 509.22 words (*SD*=255.71) and 27.44 sentences (*SD*=14.16). The texts covered a range of topics, including, as examples, scientific concepts (e.g., anthropic rock), historical events and topics (e.g., naval ships used during the World War I), and topics in botany (e.g., flora of Colombia). The mean Flesch reading ease estimate of the examined texts was 45.32 (*SD*=12.29); accordingly, the mean Flesch-Kincaid grade-level estimate was 11.63 (*SD*=2.07).

Text ratings. Several ratings were collected for each of the texts. Participants' perceived interest provided an assessment of participants' feelings of interest in the content of the texts and was measured through the 10-item Perceived Interest Questionnaire (PIQ; Schraw et al., 1995). Items were administered using a 5-point Likert scale (Schraw, 1997). Administration of the items was randomized. An example item was "*I would read more about this topic in the future.*". Scores on the PIQ items were averaged to yield a composite perceived interest score. Previous research has supported the psychometric properties of the PIQ (Sperling & Dornisch, 2010). Scores on the PIQ demonstrated adequate reliability in the current study as estimated by Cronbach's alpha (α=0.93).

Participants' ease of comprehension provided an assessment of comprehensibility of the texts and was measured by the ease of comprehension subscale of the Sources of Interest Questionnaire (Schraw et al., 1995). Items were administered using a 5-point Likert scale and administration of the items was randomized. An example item was "*The text was easy to remember.*". Previous research has also supported the psychometric properties of the SIQ (Schraw et al., 1995). Scores on the ease of comprehension ratings demonstrated adequate reliability in the current study as estimated by Cronbach's alpha (α=0.85).

Participants were administered one item that assessed their familiarity with the content of the text. The item was administered using a continuous, 100-point scale ranging from 0 (*Not at all familiar*) to 100 (*Extremely familiar*), and provided a specific rating of participants' perceived knowledge of the content of the expository texts[7].

Reading behaviors and attitudes. To capture and control for participants' reading behaviors and attitudes, several questions were administered. These questions asked for

[7] Because the text familiarity rating relied on one score, an estimate of reliability for the ratings could not be computed.

information regarding the average number of books participants read in a year, participants' reported confidence in their reading ability (on a 5-point scale ranging from *Very unconfident* to *Very confident*), and participants' reported enjoyment of reading (on a 5-point scale ranging from *Not at all* to *Very much*). Similar items have been administered in existing research examining readability metrics among adults (e.g., Crossley et al., 2017).

Design and procedures. IRB approval for the current study was obtained; informed consent was obtained from all participants prior to their participation. Participants first consented to participate in the study and read an overview of the study aims and procedures. Participants were then randomly assigned three of the 90 texts. This procedure was employed to improve the reliability of obtained text ratings while balancing study length. The order of text administration was randomized. Participants read each text, provided a one-to-three sentence summary of each text, and completed ratings of each text. Participants' reading times were recorded in milliseconds. Text summaries were screened to evaluate fidelity of study completion and to further support the reliability of the text ratings examined. After reading and rating each text, participants completed the demographics and reading behaviors and attitudes questionnaire. Participants were compensated $3.01 for their completion of the study[8]. The mean study completion time was approximately 30 minutes.

Results

Preliminary analyses. To support use of a composite score representing participants' reading behaviors and attitudes, an exploratory factor analysis was first conducted on the three items administered. The analysis was conducted using SPSS. Maximum likelihood estimation was used as the extraction method with Promax rotation employed as an oblique rotational

[8] The compensation amount was determined by Prolific's pricing estimator and was based in part on estimated study length.

procedure. A one factor solution emerged with positive loadings exceeding 0.50 for the items measuring number of books read (0.59), confidence in reading ability (0.54), and enjoyment of reading (0.94). Obtained communalities exceeded 0.20 (Kline, 2005). The solution accounted for 64.55% of the variance in the items. To represent participants' reading behaviors and attitudes, factor scores were computed and saved using the Bartlett method.

[Insert Table 9 about here]

Primarily analyses. Descriptive statistics for primary variables are presented in Table 9. Scores representing participants' reading behaviors and attitudes and text familiarity as well as text reading times were entered into the first block of the hierarchical regression analysis. In the second block, text easability metrics representing concreteness, syntactic simplicity, and deep cohesion were entered. Finally, ICD scores were entered into the third and final block of the regression. Participants' perceived interest in the texts and ease of comprehension ratings for the texts were separately predicted. Summary results of the regression analyses are presented in Table 10.

[Insert Table 10 about here]

Across models, evaluation of variance inflation factor (<2.00) and tolerance (>0.40) estimates indicated that issues with multicollinearity were not present (all condition index values were also less than 15.00). The overall regression predicting participants' perceived interest was statistically significant, F (7, 82)=2.36, $p < .05$. Scores on the measure of ICD emerged as the sole significant predictor of participants' perceived interest in the texts, b=0.11, SE=0.04, t=3.09, $p < .05$. The results indicated that increases in the ICD of the texts predicted increases in participants' ratings of interest in the texts. Combined with other variables, ICD explained a total of 17% of the variance in participants' perceived interest (R^2=0.17), and explained 11% unique

variance above and beyond reading behaviors and attitudes, reading times, and the other text metrics (ΔR^2=0.11, F (1, 82)=10.93, $p < .05$).

The overall regression predicting participants' ease of comprehension ratings was also statistically significant, F (8, 82)=3.56, $p < .05$. In the first block of the regression, participants' text familiarity ratings significant and positively predicted their ease of comprehension ratings, b=0.01, SE=0.004, t=3.41, $p < .05$. Participants with higher familiarity with the text content provided higher ease of comprehension ratings of the texts. The familiarity ratings, together with participants' reading behaviors and reading times, accounted for 12% of the variance in ease of comprehension ratings. None of the text easability ratings emerged as statistically significant, ps > .05. In the final block of the regression analysis, ICD significantly predicted participants' ease of comprehension ratings, b=0.08, SE=0.04, t=2.00, $p < .05$. The results indicated that increases in the ICD of the texts predicted higher ease of comprehension ratings of the texts. Combined with other variables, ICD explained a total 23% of the variance in participants' perceived interest (R^2=0.23), and explained 5% unique variance above and beyond reading behaviors and attitudes, reading times, and the other text metrics (ΔR^2=0.05, F (1, 82)=5.01, $p < .05$).

Finally, we examined whether the association between ICD and adults' ease of comprehension ratings depended on their familiarity with the texts that were read. Specifically, we examined the moderating role of text familiarity in the relationship between ICD and ease of comprehension ratings while controlling for participants' reading time and estimates of syntactic simplicity and word concreteness for the texts. The interaction between ICD and text familiarity was statistically significant, b=0.003, SE=0.001, t=2.10, $p < .05$, and accounted for unique variance in participants' ease of comprehension ratings, ΔR^2=0.04. The findings revealed that the contribution of ICD to ease of comprehension was significant at the 50th, b=0.09, SE=0.04,

t=2.42, $p < .05$, and 84th, b=0.17, SE=0.05, t=3.27, $p < .05$, percentiles, suggesting that the predictive effect of density on comprehensibility was positive and significant when participants were more familiar with the text content. Combined with participants' reading times and the measures of text easability, ICD and text familiarity accounted for 27% of the variance in participants' ease of comprehension ratings (R^2=0.27; F(6, 83)=5.06, $p < .05$).

Discussion

The findings of Study 3 provide evidence that ICD is predictive of measures of adults' text processing. In particular, the findings suggest that ICD serves as a significant predictor of adults' reported ratings of interest in and ease of comprehension of a variety of expository texts. This finding is noted in the presence of non-significant prediction based on commonly-examined text easability metrics, including word concreteness, syntactic simplicity, and deep cohesion. Further, in this study, ICD emerged as a somewhat stronger predictor of participants' interest in expository text. Importantly, the results of this study provide evidence of differential effects of ICD on comprehensibility of expository text based on participants' familiarity with the text content, suggesting a density-facilitative effect for participants more familiar with the text content. This finding parallels previous reader-text interaction research suggesting that readers' prior knowledge and expository text processing may depend on and be promoted by key text features (Follmer & Sperling, 2018; Kintsch, 1994). Taken together, these results provide support for the utility of ICD as a text metric predictive of adults' processing of expository texts, suggesting that expository text with a higher incidence and signaling of causal text elements fostered higher ratings of interest and comprehensibility while controlling for reading time and attitudes as well as measures of text easability.

Study 4

While the results of Study 3 provide evidence for the facilitative effects of ICD on text processing, they are based primarily on interest and comprehensibility ratings. In other words, the results of Study 3 do not directly support conclusions about the potential role of ICD in readers' comprehension of expository texts. Accordingly, in Study 4, in addition to obtaining ratings and measures of text difficulty and processing from adult readers, we measured comprehension based on an established set of 10 expository texts addressing a range of science, technology, engineering, and mathematics (STEM) concepts (Authors, 2018, 2019). Specifically, we employed a within-subjects design to examine descriptive effects of texts that were both low and high in ICD on readers' difficulty, processing (as indexed by sentence-level reading times), and comprehension. We used established measures of science text comprehension, based on the texts, that assessed both lower- and higher-level comprehension, including items that tapped literal understanding as well as relational and inferential comprehension of the STEM texts (Authors, 2018, 2019). This within-subjects approach, grounded in calculation of effect size across text processing measures, allowed us to examine variation in the role of ICD in supporting readers' perceptions and comprehension of science texts.

Method

Participants. A total of 21 participants completed Study 3 (61.9% female). Participants were recruited from a subject pool at a large, Mid-Atlantic research university in the United States. Participants received course credit for their participation. Participants in this study were primarily first- or second-year undergraduate students (*M* age=19.29). Participants did not report difficulties with reading.

Texts and reading times. Ten expository science texts were used in Study 4. The texts were designed as part of a larger funded project that examined the neurocognitive mechanisms

underlying comprehension of scientific text (Authors, 2018; Authors, 2019). The texts were designed to be approximately 300 words in length and discussed varied topics, including: the human brain, mathematics (permutations and combinations), DNA, GPS, memory, Mars, penguin species, electrical circuits, supertankers, and physics (static and sliding friction). The texts varied in content structure and represented description, comparison, and problem and solution structures (Mayer, Brandt, & Bluth, 1980; Meyer & Poon, 2001). As noted in Authors (2018), the content of the texts was based on relevant websites (e.g., NASA science outreach and Wikipedia), and underwent an extensive review and revision process to ensure accuracy of content. The average length of the texts was 305 words or 31 sentences. For additional summary of these psycholinguistic measures of these texts, please see Authors (2019).

The texts were developed in part through analysis of features of easability and readability. In particular, the texts were constructed to be above average (i.e., > 50th percentile) overall in syntactic simplicity (*M* percentile=93.4%; *SD*=5.65), referential cohesion (*M* percentile=76.6%; *SD*=17.32), and readability (Flesch reading ease=64.7%; *SD*=8.46; Flesch, 1948; Graesser et al., 2004; Graesser et al., 2011)[9]. As noted in Authors (2018), the percentage of passive sentences present in the texts was low (*M* %=9.4%). To examine participants' processing of the texts, sentence-level response times (as an index of participants' reading times) were recorded in milliseconds for all texts.

Text ratings. Immediately after reading each text, participants provided judgments of text difficulty for each of the ten expository science texts. The judgments were provided on a 5-point scale ranging from *1-Easy to understand* to *5-Difficult to understand*. The ratings provided

[9] We aimed to develop science texts that were above the 50th percentile in each of these metrics. Thus, texts were not matched on these characteristics and there existed variability in the measured degree of these easability metrics on the texts constructed.

information about readers' global perceptions of text complexity for each of the texts. Reliability of the text difficulty ratings, as estimated by Cronbach's alpha, was adequate (α=0.82). Difficulty ratings were averaged for each of the texts.

Comprehension items. Participants were given 10 multiple-choice questions per text that measured their understanding of facts and information, knowledge of the relations among concepts, and inferences based on the content presented in the texts (e.g., Kobayashi, 2004). The comprehension items are established and have been used in existing research (Authors, 2018). The comprehension items were designed to have strong text coverage, tapping understanding of main ideas and concepts across sentences in each of the texts. For example (as discussed in Authors, 2018), comprehension items assessing understanding of the Mars text contained three literal, one inference, and six relations-based comprehension items that spanned ideas and concepts in 23 of the 31 sentences of the text. Reliability of the comprehension items, based on an estimate of composite reliability, was adequate (α=0.85; Kline, 2005). Comprehension item performance was averaged to produce a text-level comprehension score for each of the texts.

Participants. IRB approval for the current study was obtained; informed consent was obtained from all participants prior to their participation. Texts were presented and read one sentence at a time; during reading, sentences were presented in a single line using sentence case. Participants read each text in a self-paced manner. The order of text administration was randomized. The mean study completion time was approximately 60 minutes.

Results

To examine the role of ICD on individuals' ratings, processing, and comprehension of the expository science texts, each of the 10 texts was subjected to linguistic analysis using Coh-Metrix (paralleling the procedures used in Studies 1 and 2). A ICD formula based on the

arithmetic mean of the three measures (incidence of causal verbs and particles, incidence of noun phrases, and incidence of verb phrases) was then applied to the texts. ICD measures for the texts ranged from 186.89 to 236.38 (range=49.50). Next, using the median ICD (*Mdn*=219.83), texts were divided into low (*n*=5; brain, circuit, DNA, GPS, and penguins texts) and high (*n*=5; friction, Mars, mathematics, memory, and supertankers texts) density groups[10]. The mean ICD of the low density text group was 209.66 (*SD*=13.28) while the mean density of the high density text group was 227.39 (*SD*=6.62), resulting in a mean ICD difference of 17.73 between text groups. Examples of low density (penguins; 306 words) and high density (Mars; 310 words) texts are presented in Appendix A; the Mars text contained a greater incidence of causal verbs and particles (60.32 compared with 36.07) as well as verb phrases (225.40 compared with 170.49).

Scores on the difficulty ratings (*M*=2.95; *SD*=0.43; range=1.24 across the ten texts), sentence-level response times (*Mdn*=2551.06; range=722.89), and comprehension items (*M*=0.86; *SD*=0.07; range=0.18) were then examined based on low and high ICD groups. Descriptively, the lower the rating of text difficulty (indicating less perceived text complexity), the greater the ICD score (see Figure 2); this mean difference (0.32) indicated a medium effect of ICD on participants' text difficulty judgments (Cohen's *d*=0.77; Cohen, 1988). Conversely, participants demonstrated increased reading times when reading texts with greater ICD (see Figure 3); this median difference (114.66) indicated a small to medium effect of ICD on participants' processing speed for the expository science texts (Cliff's *d*=-0.28; Macbeth, Razumiejczyk, & Ledesma, 2011). Finally, participants' comprehension accuracy based on the texts was higher (indicating higher mean accuracy across comprehension items based on the

[10] As noted above, this approach was taken in lieu of a correlational approach based on the size of the text sample administered and the within-subjects approach taken to the examination of RCD in Study 3.

texts) after reading texts with greater ICD (see Figure 4); the obtained mean difference (0.04) indicated a medium effect of ICD on participants' difficulty judgments (Cohen's *d*=-0.60; Cohen, 1988).

[Insert Figures 2, 3, and 4 about here]

Discussion

The findings of Study 4 revealed a number of ICD effects, based on adult readers' processing and comprehension of the texts (see Crossley et al., 2017), that extend those obtained in Studies 1, 2, and 3. In particular, consistent with the findings of Study 3, the results of Study 4 suggest that readers rated texts with higher ICD – based in part on higher incidences of causal verbs and particles and verb phrases – as easier to understand. They also performed better, descriptively, on measures of comprehension based on texts with higher ICD. However, readers spent more time processing texts with higher ICD based on sentence-level response times across the texts. These descriptive and moderate effects corroborate the density-facilitative evidence obtained above and align with emerging work in the computational linguistics literature positing moderate to strong relations between syntax-related features of text and contemporary and learner-driven measures of readability (Pitler & Nenkova, 2008).

General Discussion

Successful comprehension of expository text depends on readers' ability to forge grounded and meaningful connections between ideas in text as well as with their existing knowledge. These connections promote the development and coordination of both text-based and elaborative inferences during reading and, as a result, a more coherent mental representation of text. Models of text comprehension assume a substantive role of cohesion in promoting comprehension processes, including inferencing, and a sizeable amount of research has provided

evidence of cohesion effects on varied readers' comprehension of text (e.g., Follmer & Sperling, 2018; Kintsch, 1994; McNamara, Kintsch, Songer, & Kintsch, 1996). In short, characteristics of text such as cohesion are believed to allow readers to leverage and extend beyond their textbase level of text understanding to foster a coherent and representative situation model of text. The current research provides developing evidence for a concurrent role of content density in supporting readers' processing of expository text. While additional research is needed to further understand the density-facilitative effects suggested in this work, we offer the following interpretations and implications grounded in the construction-integration model of text comprehension (Kintsch, 2013; Pearson & Cervetti, 2017).

First, the measure of ICD examined in this work – representing the degree of causal ideas, actions, and events in expository text – appears to serve an important role in facilitating text processing. That is, based on the text metrics examined, we position ICD as a component-based metric that signals and conveys causal elaborations and relations in expository text. For example, expository text containing a greater incidence of verb phrase and causal verb information, while being more informationally compact, is likely to contain additional descriptive information regarding events, actions, and processes that are causally linked in expository text. These linkages, on the whole, may contribute to a reader's representation of text that is more connected and related. In many ways and as noted by others (e.g., Pitler & Nenkova, 2008), discourse relations are believed to be instrumental in supporting the coherence of a reader's mental representation of text. Thus, ICD may be positioned as a component-based text metric that mediates the contribution of discourse relations to text coherence. Further work is needed to better understand this possible mediating effect as well as the possible contributions of ICD to readers' mental representations of text.

Next, ICD may serve as an important dimension of the textbase during text processing. As discussed, overlap (i.e., among arguments in text) has thus far been conceived as a major driving force behind the connecting of ideas in text. The current work suggests that the presence of subordinate and, perhaps, embedded causal propositions in text (see McNamara & Magliano, 2009) may support readers' processing of expository text. This finding also bears indirectly on our understanding of the microstructure of a reader's representation of text – that is, on the influence of the ICD of text content on readers' propositional representation of text. It may be that expository text with greater ICD supports readers' generation of local inferences such that text that is denser with respect to verb phrases and causal content is likely to contain elaboration (e.g., of main points, events, and actions), providing additional ideas and concepts on which to forge connections during reading. These connections may also be enhanced by readers' prior knowledge of or familiarity with the expository text being read (e.g., Kintsch 1994). Future research is needed to more directly understand the role of ICD in facilitating causal inference generation during reading of expository text (see also Graesser & Bertus, 1998).

On the other hand, increases in the density of content, particularly as it applies to well-constructed text, may also impart perceptions of content specificity and command. More broadly, it may be that text with greater ICD provides readers with more opportunity to engage with the text content while reading. Despite being exposed to a number of STEM texts ranging in the degree of ICD, readers in Study 4 did not rate texts with higher ICD scores to be more difficult to process, nor did readers in Study 3 find naturally-occurring expository texts with higher ICD to be more difficult to understand. Our account of these findings is predicated on the notion that ICD, giving way to content specificity and elaboration, induced readers to activate their prior knowledge so as to engage more effortfully with the texts (McNamara & Magliano, 2009). Thus,

while the high-density texts may have presented more syntactic information to process, readers may not have translated this density into perceived complexity in their text ratings.

The findings of this research also complement and, in some ways, extend recent research examining text processing and familiarity in adults. For example, Crossley and colleagues (2017) obtained crowdsourced ratings of a corpus of 150 news articles, collecting text judgment information based on processing, comprehension, and familiarity judgments. They also evaluated the texts on a number of salient linguistic features, including cohesion, lexical sophistication, and measures of readability. They obtained evidence that the incidence (i.e., average number) of verbs and verb phrases, as well as noun-based information, correlated positively and moderately with text comprehension and, to a lesser degree, text processing. In addition to this work, research by Schwarm and Ostendorf (2005) and Pitler and Nenkova (2008) found that features of text density, such as average number of noun and verb phrases per sentence were correlated with and predictive of novel ratings of text readability. Thus, our findings are consistent with the work of Pitler and Nenkova to suggest an important and supportive role of causal ICD in adult readers' processing of text.

Limitations and Future Directions

This work had several limitations. First and as described, this research was applied to expository text. As a result, the extent to which ICD applies to and is instrumental in the processing of narrative and other types of text is not yet known. Second, this work leveraged a large, volunteer-driven corpus of expository texts. While this corpus allowed us to evaluate texts that varied in veracity and focused on a range of topics and concepts, additional selections of expository texts, for example cause and effect versus compare and contrast forms, may provide additional insight into the structure of ICD and its utility in accounting for readers' text

processing. Third, this research relied on a rather limited examination of text comprehension, based on the number of texts examined and the manner in which comprehension was assessed. Future work examining the possible facilitative role of ICD should incorporate additional approaches to assessing text comprehension. Fourth, this research examined adult readers' processing of expository text as well as texts that were written in English and assessed largely at the secondary to post-secondary level according to traditional, grade level measures of readability. Thus, the extent to which ICD is facilitative of younger, developing, and English language learners' processing and comprehension of text is not yet clear. Additional work is also needed in these areas to better understand the applicability and robustness of density-facilitative effects on text processing.

Conclusion

This research investigated a developed measure of expository ICD and its ability to account for indicators and actual measures of expository text processing. After some refinement to the measure, evidence across Studies 2 and 3 suggested that this ICD measure predicted and explained unique variance in estimates of text readability and easability as well as in measures of adult readers' ratings of interest and ease of comprehension based on a range of expository texts. Further, Study 4 documented descriptive effects of ICD on adults' processing and comprehension of constructed science texts. Across studies, evidence of an intrinsic text content density-facilitative effect was obtained; that is, this research suggested that texts with higher ICD may support readers' judgments and processing of expository text.

References

Alterman, R., & Bookman, L. A. (1990). Some computational experiments in summarization. *Discourse Processes, 13*(2), 143-174.

Authors (2018).

Authors (2019).

Authors (2020).

Bartunov, S., Kondrashkin, D., Osokin, A., & Vetrov, D. (2016, May). Breaking sticks and ambiguities with adaptive skip-gram. In *Artificial Intelligence and Statistics* (pp. 130-138).

Begeny, J. C., & Greene, D. J. (2014). Can readability formulas be used to successfully gauge difficulty of reading materials?. Psychology in the Schools, 51(2), 198-215.

Belanger, J. F. (1978). Calculating the syntactic density score: A mathematical problem. *Research in the Teaching of English*, *12*(2), 149-153.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135-146.

Britton, B. K., & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, *83*, 329–345.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). New York, NY: Routledge.

Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, *23*(1), 84-101.

Crossley, S. A., Skalicky, S., Dascalu, M., McNamara, D. S., & Kyle, K. (2017). Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas. *Discourse Processes*, *54*(5-6), 340-359.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Thousand Oaks, CA: SAGE Publications Inc.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221-233.

Follmer, D. J. & Sperling, R. A. (2018). Interactions between reader and text: Contributions of cognitive processes, strategy use, and text cohesion to comprehension of expository science text. *Learning and Individual Differences*, *67*, 177-187.

Golub, L. S., & Kidder, C. (1974). Syntactic density and the computer. *Elementary English*, *51*(8), 1128-1131.

Graesser, A. C., & Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. *Scientific Studies of Reading*, *2*(3), 247-269.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*, 371-398.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223-234.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, *36*(2), 193-202.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of memory and language*, *63*(3), 259-273.

Huang, E. H., Socher, R., Manning, C. D., & Ng, A. Y. (2012, July). Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1* (pp. 873-882). Association for Computational Linguistics.

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, *103*, 114-126.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability Formulas: (Automated readability index, fog count and Flesch Reading Ease Formula) for Navy enlisted personnel*. (No. RBR–8-75). Naval Technical Training Command, Millington, TN: Research Branch.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction integration model. *Discourse Processes, 16,* 193-202.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294–303.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W. (2013). Revisiting the construction-integration model of text comprehension and its implications for instruction. In D. Alvermann, N. Unrau, & R. Ruddall (Eds.), *Theoretical models and processes of reading*. (6th ed., pp. 807-839). Newark, DE: International Reading Association.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production.

Psychological Review, *85*(5), 363.

Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA, US: Sage Publications, Inc.

Kobayashi, M. (2004). Reading comprehension assessment: From text perspectives. *Scientific Approaches to Language, 3*, 129-157.

Louwerse, M. M. (2001). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics, 12,* 291-315. Mahwah, NJ: Erlbaum.

Macbeth, G., Razumiejczyk, E., & Ledesma, R. D. (2011). Cliff's Delta Calculator: A non-parametric effect size program for two groups of observations. *Universitas Psychologica*, *10*(2), 545-555.

McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, *24*, 459–488.

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.

McNamara, D. S., & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (vol.51, pp. 297–384). New York, NY, US: Elsevier Science.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of the top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16,* 72-103.

Meyer, B. J. F., & Poon, L. W. (2001). Effects of structure strategy training and signaling on recall of text. *Journal of Educational Psychology, 93,* 141–159.

Miller, J. R., & Kintsch, W. (1980). Readability and recall of short prose passages: A theoretical analysis. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(4), 335.

Muijs, D. (2011). *Doing quantitative research in education with SPSS* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus User's Guide: Statistical Analysis with Latent Variables: User's Guide*. Retrieved from: https://www.statmodel.com/html_ug.shtml.

Muthén, B. O., & Muthén, L. K. (2017). Software Mplus Version 8.

Pearson, P. D. & Cervetti, G. N. (2017). The roots of reading comprehension instruction. In S. E. Israel (Ed.), *Handbook of research on reading comprehension* (2nd ed., pp. 12-56). New York, NY: Guilford Press.

Pitler, E., & Nenkova, A. (2008, October). Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 186-195). Association for Computational Linguistics.

Ravid, D., & Berman, R. A. (2006). Information density in the development of spoken and written narratives in English and Hebrew. *Discourse Processes*, *41*(2), 117-149.

Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading*, *3*(3), 257-279.

Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of situational interest. *Journal of Reading Behavior*, *27*(1), 1-17.

Schwarm, S. E., & Ostendorf, M. (2005, June). Reading level assessment using support vector machines and statistical language models. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 523-530). Association for Computational Linguistics.

Shaoul, C. & Westbury C. (2010) *The Westbury Lab Wikipedia Corpus*. Edmonton, AB: University of Alberta. available from http://www.psych.ualberta.ca/~westburylab/downloads/westburylab.wikicorp.download.html

Sperling, R. A., & Dornisch, M. M. (2004, July). *Facilitating higher-order processing while learning from text: Effective and less effective strategy prompts*. Paper presentation at the Annual Convention of the American Psychological Association, Honolulu, HI.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Vidal-Abarca, E., Martínez, G., & Gilabert, R. (2000). Two procedures to improve instructional text: Effects on memory and learning. *Journal of Educational Psychology*, *92*(1), 107.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995b). Dimensions of situation-model construction in narrative comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 386–397.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162-185.

Table 1

Study 1: Descriptive Statistics Based on Length and Concreteness for Text Sample 1 (N=150)

Text metrics	*M*	*SD*	%ile [25th, 50th, 75th]
Word length	805.13	743.37	[387.75, 592.00, 910.25]
Sentence length	41.16	32.64	[22.00, 32.00, 48.25]
M # sentences, paragraph	2.53	0.66	[2.10, 2.50, 2.84]
Word concreteness[a]	0.90	0.82	[0.46, 0.91, 1.38]
Flesch-Kincaid GL	11.23	2.00	[9.91, 11.14, 12.47]

Note. *M*=mean. *M* # sentences, paragraph=mean number of sentences per paragraph and is presented because paragraph length has demonstrated effects on cohesion and other indexes of text easability. GL=grade level.

[a]Word concreteness statistics are based on word concreteness easability *z*-scores for each text.

Table 2

Study 1: Descriptive Statistics Based on Causal Density Metrics

Text metrics	*M*	*Mdn*	*SD*	%ile [25th, 50th, 75th]
Incidence of causal connectives	22.08	20.56	7.93	[17.03, 20.56, 26.53]
Incidence of causal verbs, particles	31.94	30.36	8.63	[26.35, 30.36, 36.51]
Incidence of noun phrases	385.07	385.42	29.69	[365.97, 385.42, 401.84]
Incidence of verb phrases	150.22	146.63	32.17	[128.75, 146.63, 165.58]

Note. Scores on each of the text metrics were, in general, not correlated with text length as indexed by word length, mean r=0.08, $p > .05$.

Table 3

Study 1: Factor Loading and Communality Information for the EFA Conducted in Study 1

Factor: Causal Density	Factor Loadings (*SE*)	Communalities
Incidence of causal connectives	0.35 (0.16)	0.12
Incidence of causal verbs, particles	0.47 (0.17)	0.22
Incidence of noun phrases	-0.52 (0.12)	0.27
Incidence of verb phrases	0.64 (0.13)	0.41
Eigenvalue	1.75	
% Explained Variance	43.63	

Note. The loading for incidence of noun phrases was negative in the presence of scores representing both incidence of causal verbs and participles and incidence of verb phrases. Negative loadings representing select text metrics are consistent with existing factor analytic work (see Graesser, McNamara, & Kulikowich, 2011). Given this finding, estimates of internal consistency reliability and corrected item-total correlations are not presented. Kline (2005). *SE*=standard error and reflects the standard estimates for the Geomin rotated loadings. All loadings were significant at the $p < .05$ level.

Table 4

Study 1: Correlations Among Causal Density and Measures of Text Readability and Easability (N=150)

	1	2	3	4	5	6
1. Causal density scores	-					
2. Syntactic simplicity	0.47*	-				
3. Referential cohesion	-0.02	-0.27*	-			
4. Deep cohesion	0.58*	0.37*	-0.11	-		
5. Flesch-Kincaid grade level	-0.33*	-0.44*	0.13	-0.11	-	
6. Flesch reading ease	0.30*	0.18*	0.01	0.13	-0.92*	-

Note. Partial correlations were computed controlling for text concreteness. Two-tailed tests were conducted.

*$p < .05$

Table 5

Study 2: Descriptive Statistics Based on Length and Concreteness for Text Sample 2 (N=150) and Whole Text Sample (N=300)

Length text metrics: Text sample 2	*M*	*SD*	%ile [25th, 50th, 75th]
Word length	720.34	612.70	[331.25, 509.50, 795.25]
Sentence length	38.64	34.02	[20.00, 27.00, 40.00]
M # sentences, paragraph	2.61	0.82	[2.05, 2.47, 2.94]
Word concreteness[a]	0.81	0.91	[0.25, 0.94, 1.31]
Flesch-Kincaid GL	11.53	2.09	[10.12, 11.50, 12.94]
Length text metrics: Whole text sample	*M*	*SD*	%ile [25th, 50th, 75th]
Word length	762.73	681.36	[361.50, 536.50, 871.50]
Sentence length	39.90	33.30	[21.00, 30.00, 46.00]
M # sentences, paragraph	2.57	0.74	[2.09, 2.49, 2.90]
Word concretenessa	0.85	0.87	[0.33, 0.93, 1.34]
Flesch-Kincaid GL	11.38	2.05	[9.98, 11.31, 12.70]

Note. *M*=mean. *M* # sentences, paragraph=mean number of sentences per paragraph and is presented because paragraph length has demonstrated effects on cohesion and other indexed of text easability. GL=grade level.

[a]Word concreteness statistics are based on word concreteness easability *z*-scores for each text.

Table 6

Study 2: Descriptive Statistics Based on Causal Density Metrics for Text Sample 2 (N=150) and Whole Text Sample (N=300)

Text Sample 2: Causal density text metrics	*M*	*Mdn*	*SD*	%ile [25th, 50th, 75th]
Incidence of causal connectives	27.57	21.40	20.60	[16.26, 21.40, 31.10]
Incidence of causal verbs, particles	32.05	32.11	8.88	[26.73, 32.11, 36.69]
Incidence of noun phrases	390.25	388.03	35.50	[363.06, 388.03, 414.49]
Incidence of verb phrases	153.35	151.51	32.76	[129.03, 151.51, 174.51]
Whole text sample: Causal density text metrics	*M*	*Mdn*	*SD*	%ile [25th, 50th, 75th]
Incidence of causal verbs, particles	32.00	31.47	8.74	[26.54, 31.47, 36.61]
Incidence of noun phrases	387.66	385.79	32.77	[364.39, 385.79, 407.00]
Incidence of verb phrases	151.78	149.29	32.44	[128.97, 149.29, 171.24]

Note. Incidence of causal connectives was trimmed from the final model.

Table 7

Study 2: Causal Density as a Predictor of Indicators of Expository Text Processing

	Predicting text easability					
	Syntactic simplicity		Referential cohesion		Deep cohesion	
	B (*SE*)	*pr*	*B* (*SE*)	*pr*	*B* (*SE*)	*pr*
Step 1						
Word concreteness	-0.11** (0.04)	-0.18	0.07 (0.05)	0.08	-0.07 (0.01)	-0.08
R^2	0.03		0.007		0.01	
Step 2						
Causal density scores	0.07*** (0.01)	0.38	-0.01 (0.02)	-0.06	0.11*** (0.02)	0.37
ΔR^2	0.14		0.003		0.13	
F	30.33***		1.43		25.02***	
Total R^2	0.17		0.01		0.14	

	Predicting readability			
	Flesch Reading Ease		Flesch-Kincaid Grade Level Estimate	
	B (*SE*)	*pr*	*B* (*SE*)	*pr*
Step 1				
Word concreteness	4.76*** (0.74)	0.35	-0.59*** (0.13)	-0.25
R^2	0.12		0.06	
Step 2				
Causal density scores	1.29*** (0.22)	0.32	-0.22*** (0.04)	-0.31
ΔR^2	0.09		0.09	
F	40.33***		26.27***	
Total R^2	0.21		0.15	

Note. Word concreteness is based on the word concreteness easability *z*-score. Syntactic simplicity, referential cohesion, and deep cohesion are likewise represented by text easability *z*-scores. *B* reflects the unstandardized regression coefficient. *pr*=partial correlation coefficient.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 8

Study 2: Causal Density and Text Easability Metrics as Predictors of Readability

	Flesch Reading Ease		
	B (*SE*)		*pr*
Step 1			
Word concreteness	4.76*** (0.74)		0.44
R^2		0.12	
Step 2			
Syntactic simplicity	3.76*** (1.26)		0.10
Deep cohesion	1.32** (0.84)		0.02
ΔR^2		0.05	
Step 3			
Causal density scores	1.11*** (0.25)		0.25
ΔR^2		0.05	
F		21.03***	
Total R^2		0.22	

Note. Word concreteness is based on the word concreteness easability *z*-score. Syntactic simplicity and deep cohesion are likewise represented by text easability *z*-scores. *B* reflects the unstandardized regression coefficient. *pr*=partial correlation coefficient.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 9

Study 3: Descriptive Statistics for Primary Study Variables

Texts, Ratings, and Behaviors	*M*	*SD*	Range
Reading behaviors and attitudes			
Number of books	12.27	14.49	60.00
Reading confidence	4.39	0.86	4.00
Reading enjoyment	3.89	1.02	4.00
Text metrics			
Narrativity	-1.23	0.39	2.01
Word concreteness	0.77	0.85	4.00
Syntactic simplicity	-0.003	0.61	3.27
Deep cohesion	-0.06	0.79	4.71
Causal density	4.12	2.93	13.00
Text ratings			
Perceived interest	2.67	0.80	3.38
Ease of comprehension	2.76	0.88	3.25

Note. *M*=mean. *SD*=standard deviation. Number of books is based on participants' report of the average number of books they read in a year. Reading confidence is based on a 5-point scale ranging from *Very unconfident* to *Very confident*; reading enjoyment is based on a 5-point scale ranging from *Not at all* to *Very much*. Text metrics are presented in *z*-scores (easability) and computed factor scores (relative content density).

Table 10

Study 3: Causal Density and Text Easability Metrics as Predictors of Adults' Text Ratings

	Perceived Interest		Ease of Comprehension	
	B (*SE*)	*pr*	*B* (*SE*)	*pr*
Step 1				
Text familiarity rating	0.01 (0.004)	0.14	0.01*** (0.004)	0.33
Reading behaviors/attitudes	0.09 (0.11)	0.09	-0.05 (0.12)	-0.02
Reading time	-0.001 (0.001)	-0.04	0.001 (0.001)	0.08
R^2	0.05		0.12*	
Step 2				
Word concreteness	0.07 (0.11)	0.08	0.19 (0.11)	0.19
Syntactic simplicity	0.04 (0.17)	0.03	-0.19 (0.18)	-0.12
Deep cohesion	0.04 (0.12)	0.04	0.11 (0.13)	0.10
ΔR^2	0.01		0.06	
Step 3				
Causal density scores	0.11** (0.04)	0.32	0.08* (0.04)	0.22
ΔR^2	0.11**		0.05*	
F	2.36*		3.56**	
Total R^2	0.17		0.23	

Note. *B* reflects the unstandardized regression coefficient. *pr*=partial correlation coefficient.
*$p < .05$. **$p < .01$. ***$p < .001$.

Figure 1

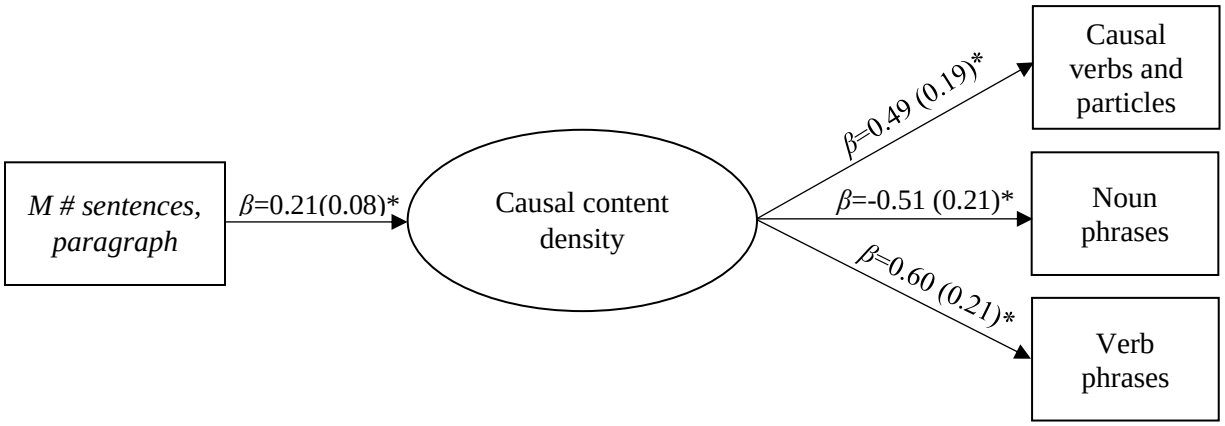

Figure 1. Confirmatory factor analysis depicting the final factor structure, after refinement, of causal density (*N*=150 texts). The latent causal density factor was regressed on the mean number of sentences per paragraph (as a measure of text length) to adjust computed factor scores. $\chi^2(1)=0.73$, $p=.39$, CFI=1.00, TLI=1.02, RMSEA=0.001, SRMR=0.01; the ratio of χ^2 to *df* was 0.73. Standardized coefficients are reported; standard errors are presented in parentheses.
*$p < .05$

Figure 2

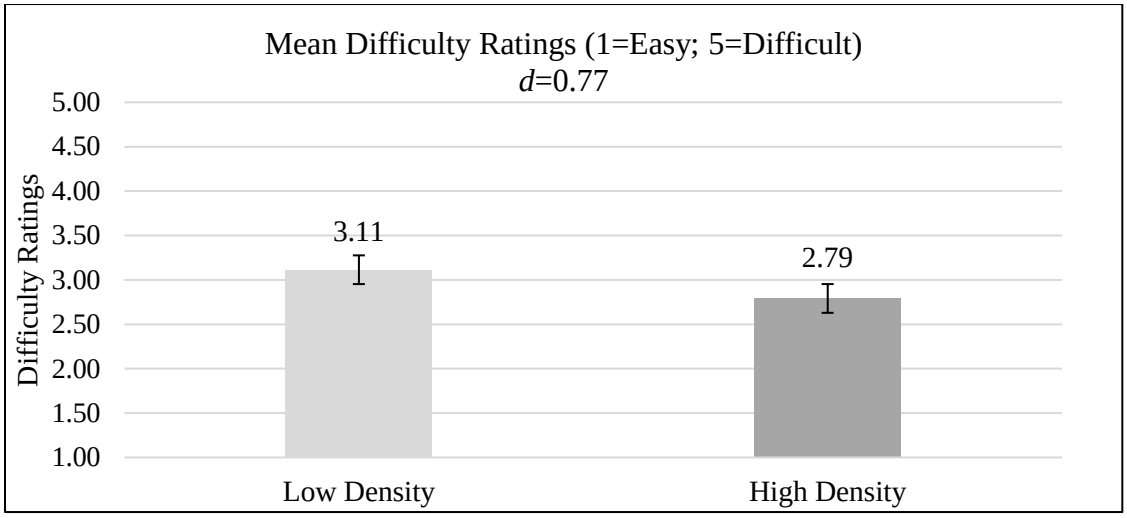

Figure 2. Mean difficulty ratings (based on a scale ranging from *1-Easy to understand* to *5-Difficult to understand*) for texts low and high in causal density. Cohen's *d* is presented as a measure of effect size, indicating a medium effect of causal density on participants' difficulty judgments. Standard error bars are presented.

Figure 3

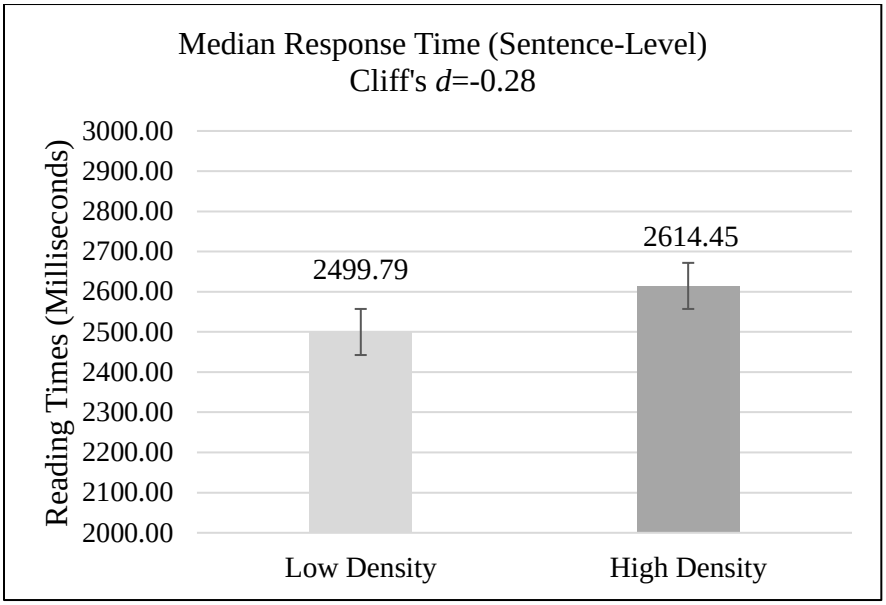

Figure 3. Median sentence-level reading times (depicted on an interval ranging from 2000.00-3000.00 milliseconds or 2-3 seconds) for texts low and high in causal density. Cliff's *d* (appropriate for ordinal or non-normal data such as those based on response times; see Macbeth, Razumiejczyk, & Ledesma, 2011) is presented as a measure of effect size, indicating a small to medium effect of causal density on participants' reading times. Standard error bars are presented.

Figure 4

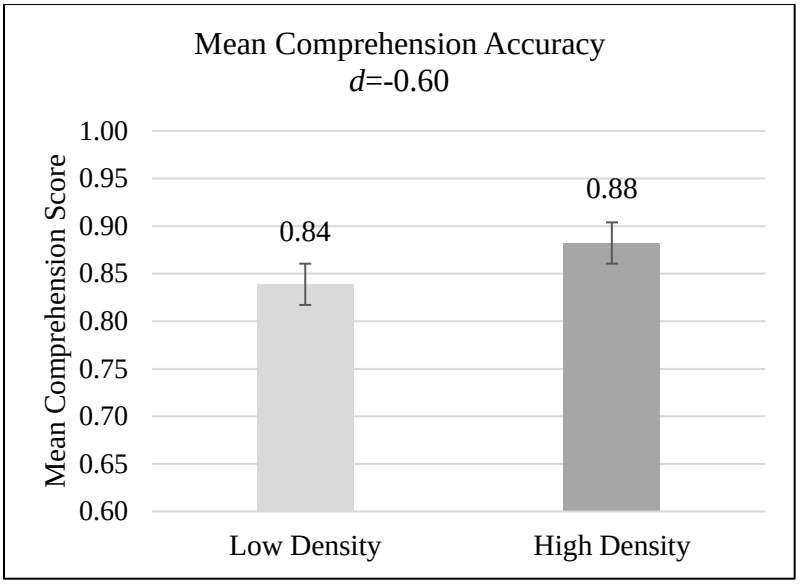

Figure 4. Mean comprehension item accuracy (ranging from 0.00 to 1.00) for texts low and high in causal density. Cohen's *d* is presented as a measure of effect size, indicating a medium effect of causal density on participants' comprehension accuracy. Standard error bars are presented.