Published in final edited form as:

Sci Stud Read. 2015 ; 19(4): 253–272.

The Relation of Knowledge-Text Integration Processes and Reading Comprehension in 7th- to 12th-Grade Students

Marcia A. Barnes,

University of Texas at Austin

Yusra Ahmed,

University of Houston

Amy Barth, and

University of Missouri

David J. Francis

University of Houston

Abstract

The integration of knowledge during reading was tested in 1,109 secondary school students. Reading times for the second sentence in a pair (Jane's headache went away) were compared in conditions where the first sentence was either causally or temporally related to the first sentence (Jane took an aspirin vs. Jane looked for an aspirin). Mixed-effects explanatory item response models revealed that at higher comprehension levels, sentences were read more quickly in the causal condition. There were no condition-related reading time differences at lower comprehension levels. This interaction held with comprehension- and inference-related factors (working memory, word and world knowledge, and word reading efficiency) in the models. Less skilled comprehenders have difficulty in knowledge-text integration processes that facilitate sentence processing during reading.

Comprehension involves the construction of a coherent mental representation of the situation described by the text (Gernsbacher, 1997; Kintsch, 1988). These situational representations connect text with the real-world situations that the text describes and, as such, may contain information about character's motivations and goals, their spatial locations, the temporal nature of events, and physical causal relations (Zwaan, Langston, & Graesser, 1995). The conditions under which integration of knowledge during reading is more or less likely to occur has been a topic of considerable research (review in Schmalhofer, McDaniel, & Keefe, 2002). Knowledge is more likely to be integrated with text to maintain local causal coherence than to elaborate on text or make predictions (McKoon & Ratcliff, 1992; Schmalhofer et al., 2002; van den Broek, Rapp, & Kendeou, 2005).

Using event-based conceptual knowledge, local coherence is routinely and rapidly maintained during reading by making knowledge-based causal bridging inferences (Matsuki et al., 2011; Singer, 2013) based on the causal relations between sentences rather than on the semantic relations between the words in sentences (Wolfe, Magliano, & Larsen, 2005). For example, fluent adult readers are faster to read the second sentence in a sequence (e.g., *Her headache went away*) and answer a subsequent question (e.g., *Does aspirin relieve headaches?*) after a stronger causal antecedent (e.g., *Jane took an aspirin*) versus after a weaker causal or temporal antecedent (e.g., *Jane looked for an aspirin*; Singer, Halldorson, Lear, & Andrusiak, 1992). In the causal condition, readers are thought to more quickly infer that “Jane’s headache went away because she took an aspirin.” Findings such as these suggest that sentence processing is facilitated to the extent that each sentence being read can be easily integrated into the prior context and that the degree of causal relatedness affects these integration processes, with stronger causal relations leading to faster reading times (Myers, Shinjo, & Duffy, 1987). For example, larger causal gaps between sentences (as in the temporal condition in the aspirin example) likely require more time-consuming elaborative processing (Myers et al., 1987) than does gap filling across shorter causal distances (as in the causal condition in the aspirin example). Although inferences are routinely made to bridge causal gaps between sentences (e.g., *The boy pushed the vase off the shelf. The floor was covered with glass.*), it is also important to note that text processing also involves integration of sentences where no inference is needed (e.g., *The boy pushed the vase off the shelf. It fell to the floor.*).

Difficulties in the ability to quickly make causal bridging inferences during reading would be expected to interfere with maintaining local coherence and constructing an accurate situation model. What is known about individual differences in inference making in general, and in knowledge-text integration processes in particular?

Children who are skilled decoders but who struggle with comprehension have difficulty integrating knowledge and text even when they can refer back to the text to answer the inference question (Cain & Oakhill, 1999), and even when the knowledge on which the inference relies is available for use (Cain, Oakhill, Barnes, & Bryant, 2001). Among college students, those who are better comprehenders are more likely than their less skilled peers to base their standards of coherence on causal relatedness than semantic relatedness between sentences (Todaro, Millis, & Dandotkar, 2010). Studies using think-aloud procedures with adults and children find that less skilled comprehenders make fewer explanatory or causal connections required for maintaining coherence than do more skilled comprehenders (e.g., Laing & Kamhi, 2003; Magliano & Millis, 2003) and that the greater the distance in the text across which two sentences need to be integrated to maintain coherence particularly affects younger children (van den Broek, 1989) and less skilled comprehenders (Cain, Oakhill, & Lemon, 2004; Magliano & Millis, 2003). These studies suggest that less skilled comprehenders are less likely than more skilled comprehenders to make causal inferences whether those inferences require the integration of explicit elements in the text or the use of knowledge to bridge causal gaps between sentences.

Most of these individual difference studies of inference and comprehension use off-line measures such as reading followed by question answering, and so assess the products of

comprehension. Rapp, van den Broek, McMaster, Kendeou, and Espin (2007) have argued that assessments that measure the *process* of comprehension and not just its *product* are necessary to better understand individual differences. Although think-aloud methods assess online processing and their findings often converge with more indirect measures of online processing such as reading times and eye fixations (Kaakinen & Hyona, 2005; see review in Rapp et al., 2007), they require readers to be aware of and reflect on what they are reading and understanding. There are few studies that use implicit measures such as reading times to assess developmental and individual differences in inference-related comprehension processes.

One study of developmental differences in inference making using sentence reading times showed that even the youngest readers were sensitive to the need to maintain causal coherence; however, there was a larger cost in increased reading times for making these causal integrations for younger readers compared to eighth graders and adults (Casteel, 1993). One of the few individual difference studies to have used sentence reading times to assess the integrity of causal bridging inference during reading was conducted by Saldaña and Frith (2007). In this study, the time to answer a question following a causally primed sequence and the time to answer the question after a similar but noncausal sequence of events did not differ between less skilled comprehenders with autism spectrum disorder (ASD) and their typically developing peers. It took less time to answer the question in the causal versus the noncausal condition. The authors concluded that knowledge-based bridging inference processes are intact in good decoders/poor comprehenders with ASD. These behavioral findings have been replicated in an eye-tracking study with small groups of individuals with and without ASD (Sansosti, Was, Rawson, & Remaklus, 2013); however, the group with ASD had longer eye fixations and more regressions than controls, suggesting that text processing was more demanding for them.

This review shows that there is little research using implicit online measures of sentence processing to investigate individual differences in knowledge-based bridging inferences. The research that does exist suggests that skilled and less skilled comprehenders may not differ in this type of inference making. However, these studies are with small groups of individuals with ASD, and their findings may not be generalizable to less skilled comprehenders without developmental disability. Furthermore, this research has compared individual differences in bridging inferences at opposite ends of the continuum of causal relatedness (strong causal vs. noncausal sequences of text). Text processing often requires gap filling across sentences that vary in the degree to which they are related from strongly to moderately to not at all (Myers et al. 1987), and so it is important to understand whether there are individual differences in how readers use causal relatedness during reading to maintain local coherence.

In addition to a lack of research on individual differences in online processing of these knowledge-based bridging inferences, the studies of individual differences in inference making just reviewed compared relatively small groups of less and more skilled comprehenders. Such designs are useful for exploratory research to generate hypotheses about potential sources of comprehension difficulties; however, they raise methodological concerns having to do with the reliability of extreme scores, interpretability of effect sizes,

and the potential for model misspecification that arises when one assumes that the relation of two variables at the extremes is similar to that in the middle of the skill distribution (Jackson & Butterfield, 1989; Preacher, Rucker, MacCallum, & Nicewander, 2005). Regression-based approaches, when sample size permits, provide estimates of effects that may be less biased and therefore more informative (Jackson & Butterfield, 1989; Preacher et al., 2005), but these types of designs are rare in the research literature on individual differences in reading comprehension in general, and in knowledge-based bridging inferences in particular.

The goal of this study was to investigate the integrity of online knowledge-based causal bridging inferences in a large sample of adolescents with a broad range of reading comprehension skill. Using materials from Singer et al. (1992), we compared reading times for the second sentence in strong causal sequences, such as “Jane took an aspirin. Her headache went away.” to those in temporal or weaker causal sequences requiring an inference across a larger causal distance, such as “Jane looked for an aspirin. Her headache went away.” Consistent with the idea that it is important to understand individual differences in the process of comprehension (Rapp et al., 2007), we used an online implicit measure developed from the cognitive science literature on reading comprehension to investigate individual differences in knowledge-text integration processes. There is strong experimental evidence that the paradigm used in the current study largely taps memory-based knowledge-text integration processes during reading (Halldorson & Singer, 2002). To address the issue of potential model misspecification associated with the use of small groups categorized as high or low in comprehension, we used a regression-based approach to more precisely test the relation of knowledge-based causal bridging inferences to reading comprehension across a broad range of comprehension performance.

Another unique aspect of this study has to do with how we tested whether comprehension-related individual differences in knowledge-based bridging inferences could be explained by inference- and comprehension-related knowledge and skills such as word and world knowledge, reading fluency, and working memory (e.g., Cain et al., 2001; Cain, Oakhill, & Bryant, 2004; Cain, Oakhill, & Lemmon, 2004; Cromley & Azevedo, 2007; Klauda & Guthrie, 2008; Oakhill, Cain, & Bryant, 2003). The fixed effects designs and analyses common to most studies assume that the effects of textual conditions such as causal relatedness are constant across the material read within a particular condition (e.g., causal vs. temporal). Similarly, characteristics that readers bring into any testing situation (e.g., word and world knowledge, working memory, reading fluency) can affect inference making and interact with text characteristics as readers strive to establish text coherence. Some studies of inference making and comprehension experimentally or statistically control for or model these factors (e.g., Cain et al., 2001; Cromley & Azevedo, 2007) to determine the unique impact of inference making on comprehension. The novel approach taken in this study was to model both the differences between individuals and the differences between items as well as their interactions (Baayen, Davidson, & Bates, 2008) by using mixed-effects, explanatory item response models. In contrast to more commonly used fixed effects models, the use of random effects models treats texts and people as samples from a larger universe and allows for inferences about effects of text and person characteristics to generalize rather than limit inferences to the specific texts and to the values of text and

person characteristics sampled in the present study. Thus, the use of random effects models for participant and text characteristics allowed us to determine whether the relation of reading comprehension to knowledge-based bridging inferences holds even when taking into account person-level and item-level factors known to be related to both bridging inference processes and reading comprehension.

In keeping with findings from individual difference studies of inference making just reviewed, we predicted that less skilled comprehenders would not show a reading time advantage for reading the second sentence in the causal versus the temporal condition. Because there is evidence that word and world knowledge, reading fluency, and working memory are importantly related to inference making and comprehension, but do not fully account for variability in these skills (e.g., Cain et al., 2001; Cain, Oakhill, & Bryant, 2004; Cromley & Azevedo, 2007; Oakhill et al., 2003), we hypothesized that the differences between good and poor comprehenders in the integration of knowledge and text would hold even after controlling for several of the factors known to be related to inference making and reading comprehension. We considered this to be a stringent test of the hypothesis that individual differences in the integration of knowledge during reading are importantly related to variability in reading comprehension.

METHOD

Participants

Participants were 1,765 students in Grades 6 through 12 from mainstream classrooms in four school districts within the greater Houston area. Selection was based, in part, on performance on the previous year's administration of the Texas Assessment of Knowledge and Skills (TAKS), the state reading accountability test and a reliable and valid measure of reading comprehension (Cirino et al., 2013). Students were randomly selected from subgroups that met or did not meet benchmark criteria on the TAKS. We randomly selected students within each class but oversampled students with poor TAKS performance, so that 47% of the sample passed and 53% did not pass. Students were excluded from participation if their school identified them as Limited English Proficient if their reading instruction or English Language Arts instruction was provided by a Limited English Proficient teacher or if they had a significant disability (e.g., intellectual-cognitive disabilities, severe behavioral disabilities, or autism). Students who consented were screened on word decoding and general intelligence. Students who scored at or above the 20th percentile on the Woodcock–Johnson III Tests of Achievement, Letter Word Identification subtest (Woodcock, McGrew, & Mather, 2007), were eligible to continue because we wanted to have a sample whose comprehension difficulties were not primarily related to very low word reading. Students also had to have a verbal and/or fluid intelligence score at or above 70, as determined by the Kaufman Brief Intelligence Test–2 (Kaufman & Kaufman, 2004) in order to rule out intellectual disability. In total, 166 students refused consent and 411 were disqualified due to low word reading. Students who passed the screening measures ($n = 1,352$) were then tested on a larger battery. Those who completed all relevant portions of the battery took part in this study. Specifically, participants from Grade 6 ($n = 156$) were excluded because they did not receive all the measures for this study, and 87 students from Grades 7 through 12 were

excluded because they did not complete all relevant tasks. In total, 1,109 adolescent students in Grades 7 through 12 (52% male) qualified and completed the relevant tasks. Participants in Grades 7 through 12 who did and did not complete all tasks ($n = 87$) were not different on demographic variables or reading comprehension. The sample was representative of the districts' student population, with 50% Hispanic, 24% White, 22% African American, 2% Asian or Pacific Islander and less than 1% Native American or multiple ethnicities. Sixty-five percent qualified for free or reduced lunch.

Measures

Screening Measures

TAKS (Texas Education Agency, 2003): The TAKS is a group administered, criterion-referenced assessment of reading comprehension, with grade specific forms. The TAKS requires students to read both expository and narrative passages and to answer comprehension questions and is designed to assess factors such as critical thinking, use of strategies, and analysis. Internal consistency for the 2010 TAKS for Grades 7 to 12 ranges from .73 to .89 and from .87 to .89 for the 2011 TAKS.

Woodcock-Johnson III Tests of Achievement, Letter Word Identification subtest (Woodcock et al., 2007): The Letter Word Identification subtest requires children to read real words that vary in difficulty. The test is individually administered and had a mean reliability coefficient of .91 across our sample.

Reading Comprehension—The Gates MacGinitie Reading Test (GMRT)—Comprehension subtest (MacGinitie, MacGinitie, Maria, & Dreyer, 2000) is a group-administered assessment of reading comprehension requiring participants to read passages of narrative and expository text silently and answer relevant comprehension questions. This test measures the products of comprehension (Rapp et al., 2007) and is less dependent on word reading and more on oral language proficiency than other similar standardized reading tests (Cutting & Scarborough, 2006). Internal consistency reliability ranges from .91 to .93. GMRT lexile scores were used in analyses as a measure of reading comprehension. The lexile score is derived from a Rasch equation which uses sentence length (syntactic component) and word frequency (semantic component) associated with the text to provide an objective measure of reading comprehension on a continuous or developmental scale of reading ability. The reader's lexile score assesses the level of text at which the student can independently read and comprehend about 75% of the text. Lexile scores are preferred to other types of comprehension scores because they allow for the use of an equal interval scale that is not age-adjusted and more revealing of age-related factors should they exist. Note that grade and age within grade are controlled in the analyses.

Inferencing- and Comprehension-Related Covariates

Decoding efficiency: The Sight Word Reading Efficiency subtest of the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) was used to assess word reading efficiency. The TOWRE is individually administered and measures students' ability to read real words out of context quickly and accurately. The internal consistency of the word reading efficiency test exceeds .95.

GMRT–Vocabulary subtest: This group-administered test of acquired reading vocabulary requires students to select a word or phrase that means the same as the test word that is presented in a brief context that cues part of speech but provides no cues to meaning (McGinitie et al., 2000). Within sample alphas averaged .87 across all grades and ranged from .83 to .90.

Working memory: Working memory was assessed using the Numbers Reversed subtest of the Woodcock–Johnson–III Tests of Cognitive Abilities (Woodcock et al., 2007). Within sample alphas averaged .79 and ranged from a low of .75 in seventh grade to .83 in 11th grade. Students are required to repeat back in reverse order a string of digits presented orally by the examiner. The task begins with a set of trials consisting of a string of three digits each. Additional trials continue with strings getting progressively longer. The task is discontinued when the student responds incorrectly to three trials in a group. A raw score consisting of the number of trials correct was used as a measure of working memory.

World knowledge: The GMRT–Background Knowledge test is an experimenter created group-administered multiple-choice test that assesses some of the word and world knowledge deemed necessary to understand relevant GMRT reading comprehension passages and questions (e.g., Who is Shakespeare? What does “fortified” mean?). Two forms were created: one form corresponding to the GMRT Grades 7–9 Comprehension subtest and one form to the GMRT Grades 10–12 Comprehension subtest. Both forms were given to all students and administered at least 48 hr prior to doing the GMRT reading comprehension test. Cronbach’s alphas on the Grade 7–9 Form ranged from .58 to .64 for standardized scores, and from .58 to .75 for standardized scores on the Grade 10–12 Form.

Inference Task—We used an individually administered computerized task with materials provided by Murray Singer (also see Singer et al., 1992, Experiment 1). A Lenovo ThinkPad laptop with a 15-in. screen was used to administer the task using an EPrime application. The task consisted of 48 test trials per participant, with 16 experimental trials and 32 filler trials. Experimental trials are further divided into causal and temporal conditions described next. Trials were presented in the same random order for all participants with the constraint that no more than three items of the same type could occur in a row and such that each block of 24 trials comprised eight experimental items (four of each type) and 16 filler items. Six practice trials were presented prior to the test trials.

A trial consisted of two sentences followed by a question. At the beginning of each trial the word “READY” was displayed in the middle of the screen. The trial was initiated when the subject pressed the space bar after which a fixation symbol appeared on the screen for 500 ms, followed by the first sentence. The subjects were instructed to press the space bar when they were finished reading the first sentence, which after 50 ms advanced the computer screen to the second sentence in the trial. When they had finished reading the second sentence, they again pressed the space bar, which removed the second sentence and displayed the question after a delay of 2,500 ms. The subject then answered the question “yes” or “no” by pressing the appropriate key—the P key for Yes and the Q key for no. After the subject had answered the question, the next trial began 1,000 ms later with the word “READY.” Reading times were recorded for Sentence 1 and Sentence 2. Response

time and accuracy were also recorded for the response to the question, which we treated as a check on knowledge needed to make the inference.

There were two types of experimental trials with eight items per condition. One type of experimental trial represented a strong causal sequence of events (e.g., Jane took the aspirin. Her headache went away), and the other type of experimental trial represented a temporal or weaker causal sequence of events (e.g., Jane looked for an aspirin. Her headache went away). After reading both sentences, students answered a question (e.g., Do aspirins relieve headaches?), which probed the hypothetical validating fact (i.e., that aspirins relieve headaches). Students were asked to respond as quickly and as accurately as they could. Two forms were created to counterbalance materials across conditions.

In addition to the experimental trials, four types of filler trials were used. Two types of *false* trial paralleled the structure of the true causal and temporal trials. Two other types of filler trials were used to ensure that the subjects read and paid attention to the first two sentences given that questions for the experimental trials can be answered without reference to the context sentences. Filler trials were not analyzed. Examples can be found in Singer et al. (1992). In sum, each participant responded to 16 test trials, eight in the causal condition and eight in the temporal condition, as well as 32 filler trials to equate the number of “yes” and “no” responses and to ensure that participants read the details contained in both of the sentences prior to answering the question.

All reading times were converted to words read per second. Within-sample reliability coefficients (Kuder-Richardson 20) were 0.73 for reading speeds for Form A and 0.68 for Form B.

Analytic Approach

Repeated measures analysis of variance models have been used in the past to account for individual differences in inferencing skills of good and poor comprehenders, where reading groups are defined based on a cut-point on a reading comprehension measure. In contrast, we modeled the effects of reader and text characteristics on inferencing skill without dichotomizing the distribution of reading comprehension.

To facilitate discussion of the models and presentation of the results, we use the terms *trials* and *items* interchangeably in so far as we are modeling reading speeds for the second sentence in a trial. Thus, although a trial consisted of the presentation of two sentences and one question and involved recording reading times for two sentence items and response time for one question item, our focus is on the reading time for the second sentence item, making items and trials synonymous in what follows.

To determine whether reader characteristics (word decoding efficiency, vocabulary, world knowledge, reading comprehension, working memory, grade, and age within grade level) and text characteristics (causal vs. temporal) made unique contributions to inference making, mixed-effects explanatory item response models were fit to the trial by trial reading speed data using SAS PROC MIXED. This approach was used in order to account for the cross-classified structure of the data (see Figure 1). Specifically, item-level reading speeds were

cross-classified between persons and items. Furthermore, people were nested within grade and form (e.g., participants were either tested using the Form A or the Form B trials) and items were nested within condition within form (i.e., for any given person, a given item was either causal or temporal depending on the form taken by that person). Given this structure, people and items were clusters at Level 2, and item-level reading speeds were treated as observations at Level 1.

Grouping variables for people, such as grade, and items (i.e., condition) were treated as fixed factors. Form was not included in the model, because items were randomly assigned to conditions within form, participants assigned the same form received the same items in the same condition, and individuals were randomly assigned to forms. Forms are relatively balanced across grades. Thus, condition and grade effects are not confounded with form, and form can safely be treated as error in the model given the large sample size and likely small effects of form. Of importance, the present study modeled reading speeds as a function of item and person characteristics. Although accuracy data are typically utilized in item-level analyses, this study focused on reading speed due to limited variability in accuracy across items and people. Response time data were collected for 48 items from 1,109 individuals (total possible observations = 53,232).

RESULTS

Prior to analysis, data were screened for missing values, outliers, and normality. Evaluations of skewness and kurtosis, as well as visual inspections of the data, showed no significant problems. Prior to removing outliers, studentized residuals were obtained by regressing each variable on grade and age within grade in order to remove the effect of age and grade on mean performance. Outliers for the covariates (e.g., word decoding, vocabulary, etc.) were defined as observations with residuals greater than $|3.00|$ and high leverage. In addition, data for individual trials were excluded from the analysis if reaction times were extremely long or short. Specifically, unusual reaction times (less than 1,000 ms and more than 10,000 ms) were discarded. Extremely fast reaction times are likely due to participants accidentally prepushing the button without having read the material. This interpretation was informed by and is corroborated by reading times (i.e., word read per second) on these trials far exceeding subjects' normal reading times on other trials and on measures of reading fluency (i.e., TOWRE). Similarly, it's difficult to know if an extremely long response time is due to a participant's inattention to the task or reading the test material multiple times. Results were similar when the data were not trimmed and when the data were trimmed for prepuses. However, because the amount of data subject to trimming was small (viz. 0.17–1% of responses across all items for prepuses and 0.001–0.10% of responses for long response times), we report only the analyses for the data that were trimmed on both ends. Response time data were transformed to reading rate, or words read per second, to account for sentence length, and covariates were standardized. Only items with correct responses were retained. Although six cases were dropped because they had less than 60% correct responses on the task, accuracy was very high (Table 1). Sample means and standard deviations are presented in Table 1 by grade, and the distribution of reading speed is presented in Figure 2, for Sentence 2 in the causal condition.

Initially, a baseline model was fit in order to estimate the magnitude of the variance due to persons and items (Table 2). A second and third model included item characteristics (causal and temporal) and reader characteristics (word decoding efficiency, vocabulary knowledge, world knowledge, reading comprehension, working memory, grade and age), respectively. Reader and text characteristics, as well as the interaction between reading comprehension and condition, were explored in the third model, which included reading comprehension as an explanatory variable on the person side, and the full model, which included all possible explanatory variables on the person side. All models were fitted using Restricted Maximum Likelihood.

In addition to the random effects of the residual error ($\sigma^2_{\text{Residual}}$), two variances were estimated in all of the models to reflect variation in the mean responses as a function of items (σ^2_{Items}), as well as people ($\sigma^2_{\text{Persons}}$), and the drop in these parameters reflect the addition of explanatory covariates included in subsequent models. These variances were uncorrelated in the model reflecting the cross-classified nature of the data structure. As the initial models are nested in the full model, the -2 Residual Log Likelihood, Akaike's Information Criteria, Akaike's Information Criteria Corrected, and Bayesian Information Criteria model fit indices were used to compare models. These indices are a function of the log likelihood, and lower values represent a better fit to the data. As can be seen by the drop in model fit indices, the final model including all persons and items covariates provided the best fit to the data. Therefore, only parameter estimates from the full model are presented next and in Table 3.

The results showed a significant effect of inferential condition, $F(1, 15,532) = 219.57, p < .0001$,¹ with parameter estimates indicating that the causal condition was significantly associated with faster reading compared to the temporal condition ($\beta = 0.21, p < .0001$). Converting the parameter estimates in Table 3 to mean reading rates revealed faster reading rates for the causal condition ($M = 3.30, SE = 0.15$) compared to the temporal ($M = 3.09, SE = 0.14$) condition.

A significant effect was found for grade, $F(5, 1022) = 2.76, p = .01$, and age within grade was marginally significant, $F(6, 1022) = 2.12, p < .0048$. Overall, students in higher grades read faster than those in lower grades, with one exception: Reading speeds were comparable for Grades 7 and 8 ($M = 3.05, SE = 0.28$ and $M = 2.72, SE = 0.20$, respectively).

In the initial model with reading comprehension as the only person level characteristic, the effect was significant ($\beta = 0.26, p < .001$). For the model with additional reader characteristics, the following main effects were found: background knowledge (Grades 10–12 Form), $F(1, 1021) = 30.82, p < .0001$; sight word reading efficiency, $F(1, 1022) = 128.35, p < .0001$; and vocabulary, $F(1, 1022) = 128.35, p < .0001$. The effect of working memory did not reach statistical significance, $F(1, 1022) = 3.90, p < .00484$. The Grade 7–9

¹Following guidelines specified in Judd, Westfall, and Kenny (2012), the degrees of freedom for the F test were estimated using the Kenward-Roger approximation. However, note that inferences based on traditional t and F distributions might be less appropriate for mixed effects models with cross-random effects for subjects and items, and alternative methods (e.g., Bayesian estimation) exist (Baayen et al., 2008).

Form for background knowledge was not included in the model because the effect was not significant in preliminary analyses. There was no main effect of reading comprehension after controlling for all covariates, $F(1, 1023) = 0.90, p = .3432$. Of importance, the interaction between reading comprehension and inferential condition was significant, $F(1, 15,533) = 11.94, p < .0005$, such that at higher levels of reading comprehension causal items were read faster than temporal items. Conversely, at lower levels of reading comprehension, reading speed was similar in both conditions. Figure 3 depicts the interaction between reading comprehension and inferential condition for items of average difficulty level. Finally, the effect size for comprehension using the 20th percentile as a cutoff was moderate ($d = 0.52, SE = 0.05$), 95% confidence interval (CI) [0.42, 0.62], and indicated that adequate comprehenders were half a standard deviation better on inference making than struggling comprehenders. The effect size for condition ($d = 0.25, SE = 0.04$), 95% CI [0.16, 0.33], as well as the interaction ($d = 0.22, SE = 0.04$), 95% CI [0.14, 0.31], were smaller but significant.

DISCUSSION

This study of adolescent readers investigated individual differences in the integrity of online knowledge-text integration processes as a function of reading comprehension level. The large sample size allowed us to investigate these knowledge-text integration processes across the continuum of comprehension skill in order to provide a more accurate and generalizable estimate of effects. In addition, the use of random effects modeling, taking into account person and text characteristics, provided a stringent test of whether putative individual differences in the relation of inference and comprehension could be explained by other factors such as reading fluency, word and world knowledge, and working memory.

There were individual differences in knowledge-based bridging inferences that were reliably associated with reading comprehension. At higher levels of comprehension, the pattern of findings was similar to that reported in studies of skilled adult readers; specifically, more skilled comprehenders read the second sentence in a stronger causal sequence of events more quickly than they did in a temporal or weaker causal sequence of events (Singer et al., 1992). In contrast, less skilled comprehenders did not show this reading time advantage for causal versus temporal sequences. Of importance, this effect remained significant even when reading fluency, word and world knowledge, and working memory were in the model. The fact that knowledge-based bridging inference shows substantial and consistent variation between individuals who differ in comprehension ability and that these individual differences are not accounted for by other inference- and comprehension-related abilities makes this type of inference a good candidate for a skill that is intrinsic to comprehension rather than simply a cognitive correlate of comprehension (Perfetti & Adlof, 2012).

These findings stand in contrast to those for less skilled comprehenders with ASD who did not differ from typically developing peers in online knowledge-based bridging inference (Saldaña & Frith, 2007; Sansosti et al., 2013). However, the current findings are consistent with off-line studies of knowledge-based inference making and with think aloud studies showing that less skilled comprehenders are less likely to make inferences to maintain causal coherence than their more skilled peers and that they are less sensitive to causal

relations in text (reviewed in Cain & Oakhill, 2007; Magliano & Millis, 2003; Todaro et al., 2010). Disparities in findings across studies could be related to difficulties in reliably discerning reading time differences between conditions in small samples and/or to differences in materials between studies, that is, comparing causal versus noncausal sequences in the ASD studies versus comparing stronger and weaker causal sequences in the current study. Alternatively, less skilled comprehenders with ASD may not have the same problems in online knowledge-based bridging inferences as do poor comprehenders without ASD.

These individual difference findings for knowledge-text integration processes are reminiscent of those for word-to-text integration processes in which words across phrase and sentence boundaries must be integrated for optimal comprehension (Yang, Perfetti, & Schmalhofer, 2007). In these studies, individual differences in the N400 component of ERP have been used to argue that less skilled comprehenders have more difficulty in paraphrase mapping between known words across sentences than do good comprehenders who make these word-to-text integrations more easily (reviewed in Perfetti & Stafura, 2014). In combination with the findings for *word-to-text integration*, we suggest that less skilled comprehenders are also less efficient in *knowledge-to-text integration* processes. That these individual differences are related in a principled way to variations in comprehension across the distribution of comprehension performance suggests that knowledge-based bridging inferences are integral to comprehension. Indeed, one contribution of this study is to show that individual differences related to comprehension level are found even for memory-based inferences that serve to maintain local coherence. In other words, individual differences in comprehension are not simply related to making only those inferences that draw more heavily on cognitive resources (e.g., working memory and/or more strategic processing) such as those that require the reader to establish global coherence across larger text distances.

What is the cognitive explanation for less skilled comprehenders' difficulties in the efficient access and integration of knowledge with text? Even though less skilled comprehenders had the knowledge needed to make these knowledge-based bridging inferences (as measured by high accuracy on question verification), they were less able than their better comprehending peers to use causal relatedness to rapidly access and integrate this knowledge during reading. Less skilled comprehenders did not show a difference in reading times between causal and temporal sequences, and they also read sentences more slowly than more skilled comprehenders. If poor comprehenders are less sensitive to causal relatedness in text (Todaro et al., 2010), causal relations in text may not serve as sufficiently strong cues to retrieve knowledge to make what are considered to be routine memory-based inferences (Halldorson & Singer, 2002; van den Broek et al., 2005). Alternatively, slower reading times in the temporal condition for more skilled comprehenders could reflect more time-consuming processes in which they engage to establish a relation between sentences across larger causal distances (Myers et al., 1987). The current study cannot differentiate between these explanations. Regardless, longer reading times that are undifferentiated across conditions may signify that it is equally difficult for less skilled comprehenders to integrate information between sentences in both causal and temporal sequences. The lack of a reading

time advantage for the stronger causal sequences might suggest that relevant knowledge is simply not accessed by poor comprehenders even in strong causal conditions. Alternatively, poorer comprehenders may be slow in their activation or retrieval of knowledge during reading such that knowledge is not retrieved quickly enough to aid online sentence integration processes. Although the current study cannot definitively distinguish between these two hypotheses, the fact that there were no comprehension-related differences in response times to read and answer the verifying question is consistent with the latter interpretation.

Although this study does not identify the sources of individual differences in knowledge-text integration processes as a function of comprehension skill, it does provide evidence that these individual differences in inference making are *not* sufficiently explained by reading fluency, word and world knowledge, and working memory. To the extent that the type of inferences measured in this paradigm are largely memory-based rather than strategic (e.g., van den Broek et al., 2005), it is perhaps not surprising that working memory did not produce a significant main effect in the models. Working memory may be implicated to a greater extent when inferences must be made across larger chunks of text than that used in the current study (e.g., Cain, Oakhill, & Lemmon, 2004) and for those types of inferences that involve higher order reasoning rather than memory-based retrieval (Evans & Stanovich, 2013).

Although word and world knowledge did not completely account for the inference findings, they were significant in the models. The availability of word and world knowledge does not always predict whether that knowledge will be used or accessed to make inferences during comprehension. For example, even when less skilled comprehenders have the requisite knowledge, they do not always retrieve or access it in the same way as skilled comprehenders (e.g., Cain et al., 2001; Henderson, Snowling, & Clarke, 2013). It is possible that individual differences in knowledge accessibility or the speed with which knowledge can be accessed affects knowledge-text integration, particularly for less skilled comprehenders (e.g., Barnes, Dennis, & Haelele-Kalvaitis, 1996). Factors known to affect the accessibility of information in long-term memory include how knowledge is encoded and connected to other concepts in memory, with deeper levels of encoding and the establishment of multiple connections and overlap between concepts making retrieval from memory more probable and more efficient (e.g., Bjorklund, 1987; Kintsch, 1994). Testing whether difficulties in knowledge-text integration processes are related to characteristics of an individual's knowledge base might take the form of assessing depth and breadth of word and world knowledge (see Ouellette, 2006; Tannenbaum, Torgesen, & Wagner, 2006, for word knowledge) as a mediator of individual differences in knowledge-text integration.

There are several strengths of the current study. The use of online reading time measures to assess knowledge-based bridging inference during reading using paradigms developed to test cognitive models of comprehension allowed us to investigate the integrity of these processes during reading in individuals who differed in their comprehension abilities. The use of a regression-based approach permitted the assessment of knowledge-based inference processes across a broad spectrum of comprehension skill. This approach proved to be sensitive for detecting individual differences in knowledge-text integration and allowed us to

estimate a small but significant interaction effect of comprehension level and causal relatedness (with all other cognitive predictors in the model). This estimate of the relation of knowledge-text inference and comprehension skill in the current study is likely to be less biased and more generalizable to the broader population of adolescent readers (Preacher et al., 2005) than those effects obtained from studies with smaller groups categorized as skilled and less skilled comprehenders. Finally, the use of mixed-effects explanatory item response models allowed for a strong test of the hypothesis that more and less skilled comprehenders would differ specifically in knowledge-based bridging inference processes. The use of this approach showed that the individual difference findings for knowledge-based bridging inference processes held even with other person- and item-level factors in the model.

Limitations of the study include an inability to determine the sources of less skilled comprehenders' difficulties in knowledge-based bridging inference processes as just discussed. It would have been informative to have included noncausal or noninferential items in the study (see Saldaña & Frith, 2007, and the previous vase example) or to more explicitly manipulate the degree of causal relatedness (Myers et al., 1987) to more fully understand the nature of difficulties in knowledge-based bridging inferences experienced by less skilled comprehenders. Although we studied knowledge-text integration processes across a broad spectrum of comprehension skill, it should be noted that students with poor decoding were excluded from participation, and so the generalizability of the findings to the entire spectrum of adolescent readers needs to be tempered by the fact that not all types of adolescent readers were included in the sample. Finally, although we think that difficulties in retrieving knowledge to maintain local coherence during reading will have negative implications for construction of the situation model, it is important to keep in mind that the paradigm we used looked at integration processes across very short sequences of text rather than at the level of a fully articulated situational representation that is constructed during the normal course of reading. Future research might directly assess the effects of difficulties in knowledge-text integration at the local level for construction of situational level representations of text. Knowledge-based bridging inferences could also be tested using reading times for sentences in longer texts along with other measures such as eye fixations to provide more information about potential sources of individual differences (Sansosti et al., 2013).

In conclusion, the findings provide further refinement of the hypothesis that inference, and particularly knowledge-based bridging inference processes, is a critical component skill or "pressure point" of reading comprehension (Perfetti & Adlof, 2012) by showing that readers of different comprehension levels differ systematically in those knowledge-text integration skills that maintain causal text coherence. The consequences of difficulties in these routine sense-making processes (Singer, 2013) are that the situation-based representations of less skilled comprehenders will lack coherence because they are poorly integrated with preceding and subsequent sentences in the text. What is less well understood are the mechanisms that account for these difficulties in knowledge-text integration. Understanding more about the sources of these difficulties will be important for thinking about the malleability of knowledge-text integration processes for intervention (Compton, Miller, Elleman, & Steacy, 2014).

ACKNOWLEDGMENTS

We acknowledge the invaluable contributions of the middle and high school staff and students in Channelview ISD, Dickinson ISD, Galveston ISD, and Humble ISD in Texas. We thank Murray Singer for sharing his materials for this study.

FUNDING

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100013 to the University of Texas–Austin as part of the Reading for Understanding Research Initiative. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the U.S. Department of Education.

REFERENCES

- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008; 59:390–412.
- Barnes MA, Dennis M, Haefele-Kalvaitis J. The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology*. 1996; 61:216–241. [PubMed: 8636665]
- Bjorklund DF. How age changes in knowledge base contribute to the development of children's memory: An interpretive review. *Developmental Review*. 1987; 7:93–130.
- Cain K, Oakhill JV. Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*. 1999; 11:489–503.
- Cain, K.; Oakhill, J. Reading comprehension difficulties: Correlates, causes, and consequences. In: Cain, K.; Oakhill, J., editors. *Children's comprehension problems in oral and written language: A cognitive perspective*. New York, NY: Guilford; 2007. p. 41-74.
- Cain K, Oakhill JV, Barnes MA, Bryant PE. Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*. 2001; 29:850–859. [PubMed: 11716058]
- Cain K, Oakhill J, Bryant P. Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*. 2004; 96:31–42.
- Cain K, Oakhill J, Lemmon K. Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*. 2004; 96:671–681.
- Casteel MA. Effects of inference necessity and reading goal on children's inferential generation. *Developmental Psychology*. 1993; 29:346–357.
- Cirino PT, Romain MA, Barth AE, Tolar TD, Fletcher JM, Vaughn S. Reading skill components and impairments in middle school struggling readers. *Reading and Writing*. 2013; 26:1059–1086. [PubMed: 24000271]
- Compton DL, Miller AC, Elleman AM, Steacy LM. Have we forsaken reading theory in the name of "quick fix" interventions for children with reading disability? *Scientific Studies of Reading*. 2014; 18:55–73.
- Cromley JG, Azevedo R. Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*. 2007; 99:311–325.
- Cutting LE, Scarborough HS. Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*. 2006; 10:277–299.
- Evans JSB, Stanovich KE. Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*. 2013; 8:223–241. [PubMed: 26172965]
- Gernsbacher MA. Two decades of structure building. *Discourse Processes*. 1997; 23:265–304. [PubMed: 25484476]
- Halldorson M, Singer M. Inference processes: Integrating relevant knowledge and text information. *Discourse Processes*. 2002; 34:145–161.

- Henderson L, Snowling M, Clarke P. Accessing, integrating, and inhibiting word meaning in poor comprehenders. *Scientific Studies of Reading*. 2013; 17:177–198.
- Jackson NE, Butterfield EC. Reading-level-match designs: Myths and realities. *Journal of Literacy Research*. 1989; 21:387–412.
- Judd CM, Westfall J, Kenny DA. Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*. 2012; 103:54–69. [PubMed: 22612667]
- Kaakinen JK, Hyona J. Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eyetracking, and recall. *Discourse Processes*. 2005; 40:239–257.
- Kaufman, AS.; Kaufman, NL. Kaufman Brief Intelligence Test, second edition (K-BIT-2). Minneapolis, MN: Pearson Assessment; 2004.
- Kintsch W. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*. 1988; 95:163–182. [PubMed: 3375398]
- Kintsch W. Text comprehension, memory, and learning. *American Psychologist*. 1994; 49:294–303. [PubMed: 8203801]
- Klauda SL, Guthrie JT. Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*. 2008; 100:310–321.
- Laing SP, Kamhi AG. The use of think-aloud protocols to compare inferencing abilities in average and below-average readers. *Journal of Learning Disabilities*. 2002; 35:437–448.
- MacGinitie, WH.; MacGinitie, RK.; Maria, K.; Dreyer, L. Gates-MacGinitie reading tests. 4th ed. Chicago, IL: Riverside; 2000.
- Magliano JP, Millis KK. Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*. 2003; 21:251–283.
- Matsuki K, Chow T, Hare M, Elman JL, Scheepers C, McRae K. Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2011; 37:913–934.
- McKoon G, Ratcliff R. Inference during reading. *Psychological Review*. 1992; 99:440–466. [PubMed: 1502273]
- Myers JL, Shinjo M, Duffy SA. Degree of causal relatedness and memory. *Journal of Memory and Language*. 1987; 26:453–465.
- Oakhill JV, Cain K, Bryant PE. The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*. 2003; 18:443–468.
- Ouellette GP. What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*. 2006; 98:554–566.
- Perfetti, CA.; Adlof, SM. Reading comprehension: A conceptual framework from word meaning to text meaning. In: Sabatini, JP.; Albro, ER.; O'Reilly, T., editors. *Measuring up. Advances in how to assess reading ability*. Lanham, MD: Rowman & Littlefield Education; 2012. p. 3-20.
- Perfetti C, Stafura J. Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*. 2014; 18:22–37.
- Preacher KJ, Rucker DD, MacCallum RC, Nicewander WA. Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods*. 2005; 10:178–192. [PubMed: 15998176]
- Rapp DN, Broek PVD, McMaster KL, Kendeou P, Espin CA. Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*. 2007; 11:289–312.
- Saldaña D, Frith U. Do readers with autism make bridging inferences from world knowledge? *Journal of Experimental Child Psychology*. 2007; 96:310–319. [PubMed: 17196973]
- Sansosti FJ, Was C, Rawson KA, Remaklus BL. Eye movements during processing of text requiring bridging inferences in adolescents with higher functioning autism spectrum disorders: A preliminary investigation. *Research in Autism Spectrum Disorders*. 2013; 7:1535–1542.
- Schmalhofer F, McDaniel MA, Keefe D. A unified model for predictive and bridging inferences. *Discourse Processes*. 2002; 33:105–132.

- Singer M. Validation in reading comprehension. *Current Directions in Psychological Science*. 2013; 22:361–366.
- Singer M, Halldorson M, Lear JC, Andrusiak P. Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language*. 1992; 31:507–524.
- Tannenbaum KR, Torgesen JK, Wagner RK. Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*. 2006; 10:381–398.
- Texas Education Agency. Texas Assessment of Knowledge and Skills (TAKS). Austin, TX: Author; 2003.
- Todaro S, Millis K, Dandotkar S. The impact of semantic and causal relatedness and reading skill on standards of coherence. *Discourse Processes*. 2010; 47:421–446.
- Torgesen, JK.; Wagner, RK.; Rashotte, CA. TOWRE: Test of Word Reading Efficiency. Austin, TX: Pro-Ed.; 1999.
- van den Broek P. Causal reasoning and inference making in judging the importance of story statements. *Child Development*. 1989; 60:286–297. [PubMed: 2924653]
- van den Broek P, Rapp D, Kendeou K. Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*. 2005; 39:299–316.
- Wolfe MB, Magliano JP, Larsen B. Causal and semantic relatedness in discourse understanding and representation. *Discourse Processes*. 2005; 39:165–187.
- Woodcock, RW.; McGrew, K.; Mather, N. Woodcock-Johnson Battery, 3rd edition (Normative update). Itasca, IL: Riverside; 2007.
- Yang CL, Perfetti CA, Schmalhofer F. Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2007; 33:55–89.
- Zwaan RA, Langston MC, Graesser AC. The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*. 1995; 6:292–297.

FIGURE 1.

Data structure.

Note. C_1 = filler condition; C_2 = causal condition; C_3 = temporal condition; solid lines = filler condition (complete cross-classification); dashed lines = causal condition (incomplete cross-classification); dotted-dashed lines = temporal condition (incomplete cross-classification).

FIGURE 2.
Distribution of reading rate by condition.

FIGURE 3.
Interaction between reading comprehension and condition.

TABLE 1

Descriptive Statistics by Grade

	Grade 7 ^a		Grade 8 ^b		Grade 9 ^c		Grade 10 ^d		Grade 11 ^e		Grade 12 ^f	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
GM-BK (Max. 15)	7.99	2.34	8.73	2.35	9.68	2.36	10.97	2.30	10.48	2.29	11.21	2.41
GMRT-V (SS)	90.75	10.56	92.23	9.40	90.85	11.90	96.66	12.03	95.69	11.11	97.96	12.40
TWSW (SS)	96.98	9.56	95.76	9.04	94.21	9.56	92.45	10.10	89.55	9.10	88.30	9.95
WJ-NR (SS)	95.02	13.68	93.20	12.37	95.80	13.73	94.21	12.12	91.85	14.69	91.99	13.58
Sentence 1 (RT)	2.37	1.08	2.54	1.13	2.75	1.16	2.93	1.25	2.92	1.22	3.06	1.23
Sentence 2 (RT)	2.50	1.08	2.64	1.13	2.85	1.17	2.97	1.19	2.97	1.19	3.08	1.26
Question (RT)	1.96	0.87	2.11	0.94	2.29	0.98	2.49	1.05	2.47	1.05	2.54	1.07
Question (PC)	95.82	6.05	96.18	5.42	95.91	5.32	96.80	5.45	96.22	5.77	97.13	4.03
Causal	96.21	7.02	96.47	6.96	96.20	6.79	96.83	6.22	96.45	7.03	97.60	5.15
Temporal	95.43	7.84	95.88	7.28	95.62	7.20	96.77	7.14	96.00	7.51	96.66	5.92
Filler	88.52	7.96	88.70	8.52	90.23	7.91	91.82	9.21	91.93	6.34	93.09	5.91
GMRT-C (SS)	89.57	13.40	91.57	11.27	90.80	14.26	96.25	11.83	92.70	11.76	94.07	14.74
GMRT-C (Lexile)	785.03	159.46	853.18	133.62	878.02	184.56	998.40	140.67	979.43	144.77	1019.28	160.07

Note. GM-BK = Gates MacGinitie Background Knowledge number of correct items (Grades 10–12 Form); GMRT-V = Gates MacGinitie Reading Test–Vocabulary; SS = standard score or scaled score; TWSW = Test of Word Reading Efficiency Sight Word; WJ-NR = Woodcock–Johnson Numbers Reversed; Sentence 1 = first sentence of knowledge-text inference item; RT = response time to read (words/second); Sentence 2 = second sentence of knowledge-text inference item; Question = question of knowledge-text inference item; PC = percentage correct; GMRT-C = Gates MacGinitie Reading Test–Comprehension.

^a
n = 175.

^b
n = 188.

^c
n = 194.

^d
n = 209.

^e
n = 197.

^f
n = 146.

TABLE 2

Model Fit Indices

	Model Fit			Variance Components		
	-2 RLL	AIC	AICC	BIC	Person	Item Residual
Baseline model	50380.30	50386.30	50386.30	50380.30	0.60	0.24 0.87
Item covariates	50161.90	50167.90	50167.90	50161.90	0.60	0.24 0.86
Person covariates	47147.00	47153.00	47153.00	47147.00	0.38	0.25 0.87
Item and person covariate (reading comprehension)	48757.50	48763.50	48763.50	48757.50	0.48	0.25 0.86
Item and person covariates (all person covariates)	46929.70	46935.70	46935.70	46929.70	0.38	0.25 0.86

Note. -2RLL = -2 Residual Log Likelihood; AIC = Akaike's Information Criteria; AICC = Akaike's Information Criteria Corrected; BIC = Bayesian Information Criteria.

TABLE 3

Parameter Estimates for the Person and Item Covariates Models

Parameter	Reading Comprehension			All Person Covariates			Confidence Intervals (Lower, Upper)
	Fixed Effect	SE	Fixed Effect	SE	t Value	Least Squares Means (SE)	
Intercept (Temporal conditions)	2.67**	0.31	2.94**	0.28	10.22	3.09 (0.15)	2.79, 3.38
Causal condition	0.21*	0.01	0.21*	0.01	14.82	3.30 (0.15)	3.00, 3.60
Grade 7						3.04 (0.28)	2.50, 3.60
Grade 8	-0.11	0.33	-0.35	0.30	-1.16	2.72 (0.20)	2.33, 3.11
Grade 9	0.34	0.29	0.10	0.27	0.38	3.15 (0.14)	2.86, 3.43
Grade 10	0.37	0.29	0.08	0.27	0.30	3.14 (0.14)	2.84, 3.43
Grade 11	0.73*	0.33	0.26	0.30	0.84	3.32 (0.20)	2.91, 3.72
Grade 12	1.39**	0.42	0.73	0.38	1.89	3.79 (0.31)	3.18, 4.42
Age (Grade 7)	0.00	0.01	0.00	0.01	0.07		
Age (Grade 8)	-0.02	0.01	-0.02*	0.01	-2.15		
Age (Grade 9)	-0.01	0.01	0.00	0.01	0.30		
Age (Grade 10)	-0.01	0.01	0.00	0.01	-0.75		
Age (Grade 11)	-0.02*	0.01	-0.01	0.01	-1.47		
Age (Grade 12)	-0.03**	0.01	-0.02*	0.01	-2.33		
WJ numbers reversed			0.04*	0.02	1.98		
GM background knowledge			0.08**	0.03	2.72		
TOWRE sight words			0.27**	0.02	11.33		
GM vocabulary			0.18**	0.03	5.55		
GM reading comprehension	0.26**	0.03	0.01	0.03	0.17		
GM reading comprehension × Causal condition	0.05**	0.01	0.05**	0.01	3.46		

Note. WJ = Woodcock-Johnson; GM = Gates MacGinitie; TOWRE = Test of Word Reading Efficiency.

* $p < .05$.

** $p < .001$.