Inferential comprehension differences between narrative and expository texts: A systematic

review and meta-analysis

Virginia Clinton, Terrill Taylor, & Surjya Bajpayee

University of North Dakota

Mark L. Davison

University of Minnesota

Sarah E. Carlson

Georgia State University

Ben Seipel

California State University, Chico

Author Note

Please cite as the following:

Clinton, V., Taylor, T., Bajpayee, S., Seipel, B., Carlson, S.E., & Davison, M. (2020). Inferential
        comprehension differences between narrative and expository texts: A systematic review
        and meta-analysis. Reading and Writing 33, 2223–2248. https:10.1007/s11145-020-
        10044-2

Abstract

Inferential comprehension is necessary to connect ideas in a text together in a meaningful manner. There have been multiple studies on inferential comprehension involving texts of different genres (narrative and expository), but not a coherent overview of the findings of inferential comprehension by genre. The purpose of this study is to provide a coherent overview by conducting a meta-analysis of the findings of inferential comprehension by genre. A systematic review of the literature yielded 18 reports with 19 independent studies involving 1,714 participants in which inferential comprehension was compared by genre. Based on Robust Variance Estimation, scores on measures of inferential comprehension were higher for narrative texts than expository texts ($g = .36$, $p = .02$). This effect did not vary depending on whether inferential comprehension was assessed during or after reading, whether the texts for each genre were matched for readability, whether the reader was an adult or child, and whether the inference connected different ideas in the text (text connecting) or the text to background knowledge (knowledge based). Potential explanations of genre differences in inferential comprehension and future directions for research are discussed.

Keywords: genre, inference, meta-analysis, reading comprehension

Inferential Comprehension Differences Between Narrative and Expository Texts: A Systematic

Review and Meta-Analysis

Successful reading comprehension requires connecting ideas in a text together in a

meaningful manner in a process known as coherence (van den Broek & Gustafson, 1999).

Readers connect these ideas together through a process known as inference generation.

Inferences can be categorized in a number of manners, but generally either connect the text with

previously-read information in the same text or to readers' background knowledge (Cain et al.,

2004). In general, inferential comprehension is considered more challenging with expository

texts than narrative texts (Graesser et al., 2011). This is thought to be due to differences in both

content and structure by genre as narrative texts are centered around character goals whereas

expository texts are more varied in structures (Graesser et al., 2004; Lorch, 2015). In addition,

the background knowledge to generate inferences for narratives is typically based on everyday

life experiences whereas expository texts may require background knowledge that readers often

do not have, making it more challenging to generate inferences (Graesser, et al., 2004). Indeed,

there have been studies that show expository texts appear to be more challenging than narrative

texts (Denton et al., 2015; Haberlandt & Graesser, 1985). However, there have also been studies

of inferential comprehension that do not necessarily indicate that expository texts are more

difficult than narrative texts (Wolfe & Woodwyk, 2010). These conflicting findings necessitate a

comprehensive overview of what research has been conducted and what the inferential

differences by genre are.

Theoretical Frameworks

In the construction-integration model of text comprehension, texts are considered to be

represented at three different levels: the surface structure, the textbase, and the situation model

(Kintsch, 1998). The surface structure of a text consists of the words and syntax in that text and the ideas in a text, referred to as propositions, are created from these words and syntax. The textbase level of representation is developed when the propositions of a text are connected with each other. Finally, readers construct an in-depth mental representation of the text, through what is known as a situation model, by integrating their relevant background knowledge with the textbase. Each of these levels of representation is known to have differences by genre that could influence inference generation. The surface structure of expository texts is generally considered more difficult than that of narrative with words in expository texts being longer and less frequently used than words in narrative texts (McNamara et al., 2012). In terms of syntax, late elementary and middle school students, when instructed to write a narrative compared to an expository text on the same topic, have been shown to use more complex syntax, as indicated by the number of words per clause, for expository texts than for narrative texts (Beers & Nagy, 2011). This difficulty in surface structure could lead to readers struggling to make connections by generating inferences. Taken together, these findings indicate that the surface structure in expository texts is more difficult than in narrative texts. This difficulty in surface structure could explain why expository texts are more challenging to recall than narrative (Collins et al., 2019; Graesser et al., 1980).

There is also evidence that textbase representation construction could differ by genre. One method of assessing how easily a textbase representation can be constructed is through latent semantic analysis (LSA). In LSA, the conceptual overlap of words is measured based on the likelihood that those words would likely co-occur in other texts with higher LSA scores indicating greater conceptual overlap. In general, the more conceptual overlap within a text, the more cohesive the text is, thereby facilitating the development of the textbase (Crossley &

McNamara, 2012). Narrative texts have lower LSA scores than expository texts (McNamara et al., 2012), which would indicate that textbase construction could actually be more challenging in narrative than expository texts in terms of connecting the concepts across words.

Pronouns may be useful for constructing the textbase if readers can easily connect pronouns to their referents (Gernsbacher, 1989; Graesser et al., 2011). Third person singular pronouns (e.g., he, she, it) may be particularly useful in facilitating text cohesion as they indicate there is only one referent to track (Gernsbacher, 1989; Walkington et al., 2015; Walkington et al., 2019), and third person singular pronouns occur more often in narratives than in expository texts (van Hell et al., 2005). Also, narrative texts have more connectives (e.g., also, as well as, next, because) than do expository texts (McNamara et al., 2012). Connectives can assist a reader with comprehending both narrative and expository texts and they can guide the reader to connect relevant parts of a text (van Silfhout et al., 2015). Given that narrative texts may have advantages in textbase construction in terms of third person singular pronouns and connectives, but may actually be more difficult in terms of conceptual overlap, these differences may even out to similar textbase construction difficulty between the genres.

The background knowledge involved in constructing a situation model representation may differ substantially by genre. For instance, narrative texts typically involve background knowledge from everyday life experiences whereas expository texts are more esoteric (Graesser et al., 2004). Readers may lack knowledge of some of the information in a narrative, but readers may be able to compensate for this through knowledge of the overall schemas and story grammar in narratives (Graesser et al., 2004). In contrast, inferences generated during expository text comprehension may rely more on background knowledge, and if readers do not have relevant background knowledge, they may not be able to make inferences (Graesser et al., 2004). This

may be why background knowledge is more predictive of comprehension for expository texts
than for narrative texts (Best et al., 2008).

Standards of Coherence

Readers' criteria for a sufficient level of comprehension for a reading task are known as
standards of coherence with higher standards of coherence yielding more inferences (Oudega &
van den Broek, 2018). A reader's standards of coherence are thought to vary based on several
factors, such as a reader's background knowledge regarding the subject, the skill level of the
reader, the purpose for reading (e.g., to study vs. for entertainment; van den Broek et al., 2001;
Oudega & van den Broek, 2018). Text characteristics, such as genre, are factors in readers'
standards of coherence (Oudega & van den Broek, 2018). However, there are conflicting
findings on genre and standards of coherence. In one study comparing genre expectations for the
same text, readers constructed better mental representations of the text when they thought they
were reading a nonfictional news account, than when they were told they were reading a
fictional, literary story (Zwaan, 1994). This would indicate higher standards of coherence for
expository text than narrative text. Another study had similar findings in which standards of
coherence appeared to be higher when reading expository than narrative texts that were not
entirely the same, but had common sentences (Wolfe & Woodwyk, 2010). In this study, readers
made more connections to common knowledge when reading common sentences in the
expository texts than in the narrative texts (Wolfe & Woodwyk, 2010).

However, there are reasons to argue standards of coherence would be higher, as indicated
by higher inferential comprehension, for narrative texts than expository texts. Texts that have
clear structures and goals appear to elicit higher standards of coherence than texts that are less
clearly structured (Bohn-Gettler & Kendeou, 2014; Oudega & van den Broek, 2018), perhaps

because readers can more easily make inferences. Narrative texts are generally more clearly structured than expository texts (Lorch, 2015); subsequently, more inferences would be anticipated with narrative texts than with expository texts as has been noted in some studies (Denton et al., 2015; Haberlandt & Graesser, 1985; Karlsson et al., 2018). Genre expectations may vary with other features of the text. In another study on genre expectations, informing readers that the text was a fictional story or newspaper article did not affect overall reading comprehension scores (Schmitz et al., 2017). However, improvements to the global cohesion of a text (through paragraph transitions and summary statements) benefited performance when readers anticipated they were reading an expository (i.e., newspaper article) as opposed to narrative text (i.e., story; Schmitz et al., 2017). It is possible that the readers in this study needed more guidance when they read the expository texts as they may have anticipated the structure to be less clear.

Standards of coherence also fluctuate based on reader characteristics. For example, readers can be categorized based on the processes they use most often while reading, such as paraphrasing the text or appropriately making inferences (e.g., McMaster et al., 2012; Seipel et al., 2017). Most readers are in the same process category for narrative and expository texts; however, if they do use different processes they are more likely to use fewer inferences for expository texts than for narrative texts (Karlsson et al., 2018). In addition, in terms of readers' comprehension skill and their standards of coherence, high-comprehending readers have been found to adjust their reading strategies depending on genre whereas low-comprehending readers read both genres similarly (Kraal et al., 2019). This may be because low-comprehending readers struggle with making inferences in both genres whereas high-comprehending readers may make

inferences in narrative texts rather easily, but expository texts provide more of a challenge (Kraal et al., 2019).

Fluctuations in standards of coherence can be noted in readers from an early age and become more sophisticated as reading skills mature (van den Broek et al., 2012; van den Broek & Kendeou, 2017). For example, both early and older adolescent readers made appropriate inferences to establish cause and effect in narratives, but the older adolescents made inferences across abstracts concepts (e.g., the theme or message of the narrative) that younger adolescents did not make (Williams, 1993). Indeed, eight-year old children make appropriate inferences while reading and even preschool children generate inferences while listening to stories (Kendeou et al., 2008). Skilled adult readers typically do not need to exert conscious effort when making inferences to construct a representation of the text; in contrast, these inferences require considerable effort from children and struggling readers (van den Broek et al., 2009). For example, college students performed better on assessments of inferential comprehension of narratives than did seventh-grade readers (Franks, 1997). In addition, college students had better memory of stories if they answered inferential questions while reading, but school-aged readers did not benefit from answering questions while reading (van den Broek et al., 2001). Age-related changes in inference skills may vary by genre given the one study involving listening comprehension of both genres found comprehension performance for narratives improved from second to sixth grade, but there were no reliable differences in grade level for expository texts (Lehto & Anttila, 2003).

Inference Categorizations

Inferences may be categorized in a variety of taxonomies (Kendeou, 2015; Perfetti & Stafura, 2015). One method of inference categorization is based on the source of information for

the inference, and source can come from the text and the reader's background knowledge (i.e., text-based and knowledge-based inferences; Magliano et al., 1999). Text-based inferences, also known as text-connecting inferences, connect information previously read in the text with the currently read text. This information from the text may be reactivated from readers' memory or by looking back to the previous-read text (van den Broek et al., 2011). Information from the text can also be used to infer the meaning of novel words (Cain et al., 2004). In contrast, knowledge-based inferences use information from personal experiences or background knowledge about the world (Graesser et al., 1994). Knowledge-based inferences are also referred to as gap-filling inferences when readers need to fill in ideas not explicitly stated in the text with their background knowledge (Baker & Stein, 1981; Cain & Oakhill, 1999).

Inferences may also be categorized based on how they are used to construct a mental representation of the text. Bridging inferences, also referred to as connective inferences, are essential to connect explicit information in the text (Singer et al., 1992). A classic example of this are the statements "Katie poured water on the bonfire. The fire went out." In order to connect these statements in a meaningful manner, the reader needs to activate and apply the background knowledge that fire can be extinguished with water (Zwaan & Rapp, 2006). Bridging inferences may also be explanatory inferences that provide a reason for an event occurring (van den Broek et al., 2011). In contrast, elaborative inferences enhance the situation model, but are not necessarily critical for maintaining coherence. For example, consider the statements "Lexi wanted pancakes for breakfast. She ate them with syrup." Readers may include their background knowledge to develop the inference that Lexi used a fork to eat her pancakes, which is quite likely given how pancakes are typically consumed. However, such an elaborative

inference is not essential to connect these two statements. Predictions about what will happen in the text are also considered a kind of elaborative inference (Zwaan & Rapp, 2006).

Inferences may be categorized based on whether they are generated during reading, also known as online, or after reading, also known as offline (Kendeou, 2015). This difference may be relevant for inference assessment (i.e., whether inferences are measured during or after reading), but is generally not considered a critical difference in the inferences themselves (Kendeou, 2015). Inference generation may be measured during reading through think alouds and lexical decision tasks. In a think aloud, readers are instructed to articulate their thoughts as they read a text in order to examine their processes while reading (Ericsson & Simon, 1998; Pressley & Afflerbach, 1995). A lexical decision task involves having readers respond periodically to prompts to determine whether a letter string is a word in English (Singer, 2007). These words can be designed to be related to inferences a reader could have at that point in the text (e.g., Hamilton et al., 2016; Virtue et al., 2006). For example, the sentences "The college student forgot to set her alarm. She missed her early morning class." could be followed by a prompt to determine if "overslept" is a real word. If the reader had made the inference that the student missed her class because she overslept, the reader should verify that is a real word more quickly than if such an inference had not been made.

Inferences can be assessed in numerous manners after the text is read such as sentence verification tasks, multiple-choice questions, and open-ended response questions. In a sentence verification task to assess inference generation, readers are asked to indicate whether a sentence based on an inference one could deduce from the text (i.e., the information in the sentence was not explicitly stated in the text) is true or false based on the information they read in the text (Carlisle, 1989; e.g., Griffiths et al., 2016). Multiple-choice questions and open-ended response

items can also be designed to assess comprehension of ideas that are not explicitly stated in the text (e.g., Clinton & van den Broek, 2012; Margolin et al., 2013).

The Current Study

The purpose of this review is to synthesize and conduct a meta-analysis on the research findings on inferential comprehension in narrative and expository texts. To address this purpose, moderators that may interact with genre differences in inferential comprehension are examined. Given the numerous differences in surface level representation previously discussed, one important moderator to consider is if the texts of different genre were matched in some manner for surface and textbase features (which we will refer to as readability as these features affect how easily a text is read). Matching for readability would reduce the likelihood any potential differences in genre would likely be due to factors beyond surface structure and textbase differences (e.g., background knowledge or difference in standards of coherence based on genre expectations). Readability is often assessed through the length of the words, sentences, and texts as a whole with longer being more difficult to read (Collins-Thompson, 2014). For example, the Flesch-Kincaid metric based on the average number of words per sentence and the average number of syllables per word (Kincaid et al., 1975). Other commonly used measures are the complexity of the vocabulary and syntax as well (Collins-Thompson, 2014).

Another potential moderator is whether inferential comprehension was measured during or after reading. Examining this moderator would promote understanding of when in the reading process (during situation model construction or after reading) any potential genre differences in inferential comprehension occur. In addition, the age of the reader is important to consider because inferential skill generally improves with age (van den Broek, 2009). Finally, the source of the information for the inference (text or background knowledge) is examined as a moderator.

In terms of textbase representation, there are characteristics of both genres that would indicate easier construction (Graesser et al., 2011; van Hell et al., 2005; van Silfhout et al., 2015). Therefore, the lack of notable cohesion differences between the genres may carry over to potential differences in text-connecting inferences. However, the background knowledge for inferences is considered more challenging for expository texts than narrative. For this reason, knowledge-based inferential comprehension for expository texts may be more difficult than for narrative texts.

There are two research questions to guide this meta-analysis:

1) Overall, do readers tend to generate more inferences with narrative or expository texts?

2) Does inferential comprehension by genre vary depending on whether or not the texts for the different genre were matched for readability in some manner, whether the inferential measure is during or after reading, whether the reader was an adult or child, or what was the information source for the inference (text or background knowledge)?

Method

A systematic search for studies comparing inferential comprehension by genre was conducted. The inclusion criteria included the following: 1) empirical data were collected, 2) the texts used in the studies were longer than one sentence, 3) the texts used in the studies were read independently (not listened to), 4) participants had sufficient reading skills to independently read study passages and respond to inferential measures of their reading, 5) there was a measure specifically assessing inferential comprehension either while reading (e.g., think aloud) or after reading (e.g., short-answer questions), and 6) necessary statistics for the meta-analysis were reported or were provided by the author of the report when contacted.

In September 2018, searches for relevant literature were conducted using the search terms "narrative" AND "expository" AND "inference" in the databases SCOPUS, ERIC, PsychInfo, DOAJ, and Proquest. This yielded 1,281 citations and after duplicates were removed, there were 1,193 citations. Abstracts were screened based on the inclusion criteria and the full texts of 66 reports were examined for relevance. This resulted in identifying 14 relevant reports. The authors of these reports were emailed with requests to share any relevant reports they had knowledge of. This yielded one more report. In addition, the citations in the reference lists of the relevant reports were searched, which yielded four more reports, three of which had sufficient statistical information reported or the authors provided the necessary statistics for inclusion. One report included two independent studies (Zwaan, 1994). This led to a final number of 19 studies for the meta-analysis with a total of 1,714 participants.

Coding

In preparation for analysis, eligible reports were coded for descriptive purposes, to assess study quality, and to provide information for moderator analyses. The descriptive information included bibliographic information, ages of the participants, number of participants, descriptions of texts used, inferential measures, and the overall findings by genre. Study quality information included the reliability of the inferential measure and whether the texts for each genre were matched in some manner (and if so, how were they matched). The moderator information was whether the different genres were matched for readability and whether the inferential measure was collected during (e.g., a think aloud), after reading (e.g., comprehension questions), whether the readers were children or adults, and inference type. Following recommendations from Borenstein and colleagues (2009), a minimum of six effect sizes for a particular moderator category was necessary for moderator analyses to be appropriate (see Elleman, 2017, for a

similar approach). The descriptive information was coded and double coded by the first author

with a research assistant independently coding 25% (see Follmer, 2018, for similar approach; κ =

.80).

Statistical procedures

Hedges' *g*, which is a standardized mean difference metric, was calculated to aggregate

findings across studies. Hedges' *g* is appropriate for meta-analyses as it is corrected for bias due

to the size of the sample (Hedges, 1981). The Hedges' *g* for each effect size was calculated using

Comprehensive Meta-Analysis (version 3; Biostat) by entering in the descriptive statistics of the

inferential measure (e.g., means and standard deviations) or inferential statistics (e.g., t-tests),

depending on what was reported. Only valid inferences were included because invalid inferences

are not considered beneficial for comprehension (Cain et al., 2004). If necessary statistical

information was not reported, the corresponding author was contacted with a request for that

information. If necessary statistical information could not be determined, then that report was not

used in the meta-analysis. A positive Hedges' *g* indicates a higher score on the inferential

measure for the narrative texts compared to the expository texts.

There were studies in which multiple inferential measures were reported, and thus,

multiple effect sizes were calculated. These effect sizes were not independent because they were

from the same sample and aggregating multiple effect sizes within a study can yield inaccuracies

(Tanner-Smith et al., 2016). For these reasons, robust variance estimation (RVE) was used to

account for dependent effect sizes within studies (Tanner-Smith et al., 2016). To conduct RVE,

the package "robumeta" was used (Fisher & Tipton, 2014).

In order to examine the heterogeneity of the effect sizes, the I^2 index was used. The I^2

index is an estimate of the percentage of variability across studies that is assumed to be from

heterogeneity rather than chance (Higgins & Green, 2011). The I^2 index has a range of 0-100 and the lower the I^2, the less heterogeneity. Following Bloch (2014), moderator analyses were warranted if the I^2 index was about 20%.

Results

RQ1: Do readers tend to generate more inferences with narrative or expository texts?

Based on RVE analyses, assuming correlated dependent effects within studies of .8, inferential comprehension was greater for narrative texts than for expository texts based on 19 studies with 38 effect sizes, Hedges' $g = .36$, $SE = .14$, 95% CI [.07, .66], $p = .02$, see Table 2 for findings by study. There was an I^2 of 97.24 indicating substantial heterogeneity in the findings. A sensitivity test was conducted examining different levels of correlations for the dependent effects. This did not change the results (see Table 3).

Publication bias, in which only statistically significant results are reported, was examined using Egger's regression test of the intercept in which statistically significant results are considered an indication of publication bias. Based on the results of this test, there was no indication of publication bias, $\beta = 1.11$, $SE = 1.68$, 95% CI [-2.30, 4.52], $p = .27$.

RQ2: Variations of inferential comprehension by genre

A meta-regression model with three of the moderators (whether genres were matched in terms of readability, whether the inferential measure was during or after reading, and the whether the readers were children or adults) as coefficients was estimated (see Tipton & Pustejovsky, 2015). As shown in Table 4, neither of these moderators examined were significant. However, it is uncertain whether the lack of significance indicates a true lack of difference or insufficient power to detect an effect.

There was not sufficient information about the inferences to code all of the measures as text connecting or knowledge based. Therefore, the moderator of inference type (text connecting or knowledge based) could not be included in the meta-regression model with the other moderators as robumeta does not handle missing data (Polanin et al., 2017). A separate meta-regression model with only effect sizes that could be coded as text connecting or knowledge based was estimated (11 studies and 22 effect sizes). Based on this model, inference type was not a significant moderator (see Table 4).

Discussion

The purpose of this study was to synthesize and conduct a meta-analysis on the findings on inferential comprehension by genre. This purpose was examined because of the need to further understand how inferences are generated during successful reading comprehension across text types (van den Broek & Gustafson, 1999). Based on the findings of this meta-analysis of 19 studies, there was more inferential comprehension noted for narrative texts than for expository texts. Moderator analyses did not indicate that this effect varied depending on whether the texts by genre were matched for readability, whether the inferential comprehension measure occurred during or after reading, or whether the inferential measure was text connecting or knowledge based.

The findings from this meta-analysis support arguments in previous literature that inferential comprehension is more challenging with expository text than narrative text (Graesser et al., 2011; Graesser et al., 2004). One reason for this could be that narratives are structured around character goals and have a temporal structure that facilitates inferential comprehension—readers know that they should expect ideas in a text to connect to the characters goals and follow a logical timeline (Graesser et al., 2004; Lorch, 2015). In contrast, expository texts have a variety

of structures, and readers may find it challenging if they do not know how to appropriately

process the text in various structures (Meyer, 1987). This possibility is supported by meta-

analyses which have indicated that instruction on different types of expository structures

improves comprehension of expository texts (Hebert et al., 2016; Pyle et al., 2017).

In the summary of the studies in Table 1, it can be noted that a wide variety of inference

types, both in terms of the source of the information used and the necessity in establishing

coherence, were examined (e.g., text connecting, knowledge based, elaborative, predictive). It

was proposed that inferential comprehension that requires background knowledge may be

particularly challenging for expository texts compared to narrative because of the different types

of background knowledge involved in the genres (Best et al., 2008). The results from the

moderator analysis in this meta-analysis indicate that this was not the case—genre differences in

inferential comprehension were similar regardless of whether the inferences involved were text

connecting or knowledge based. The similar findings for text-connecting and knowledge-based

inferences by genre could be due to both the textbase and situation model levels of

comprehension being easier for narrative texts than for expository texts. One reason the textbase

representation of narratives could be easier to construct is because of similarities within the genre

with structure (Lorch, 2015). The construction of the situation model in narrative texts may be

easier to construct than that of expository texts because of the background knowledge necessary

is more common for narrative texts than expository texts (Best et al., 2008). It should be noted,

however, that there were limitations in conducting moderator analyses based on the number of

outcomes and it is unknown if the lack of an effect is due to a true lack of effect or merely a lack

of power (Hempel et al., 2013). Given this, the results regarding text-connecting and knowledge-

based inferences by genre should be interpreted with caution.

The surface structures of narrative texts are considered easier than that of expository texts. For example, the vocabulary in terms of word length and frequency of narrative texts is easier than expository texts (McNamara et al., 2012). For this reason, the potential moderating effect of having the texts matched for readability in some manner was considered. Based on the findings, there did not appear to be an interaction between matching for readability and genre differences in inferential comprehension. One interpretation of this could be that the surface structure characteristics are not the reason for the noted genre differences. Indeed, one study that examined reading performance broadly by genre found that performance on assessments was higher for narrative texts than expository even when controlling for sentence length, word frequency, and cohesion (Kulesz et al., 2016). However, this would not explain why three out of four studies in which the text (either critical sentences or the entire text) was identical found better inferential comprehension with expository texts than narrative texts (Bowyer-Crane, 2002; Wolfe & Woodwyk, 2010; Zwaan, 1994, Experiments 1 and 2). It may be that readability matching approaches need to consider more text characteristics in order to fully examine this issue. For example, measures of readability based on word and sentence length were frequently used for matching (e.g., Kraal et al., 2018; Margolin et al., 2013; Saenz & Fuchs, 2002). Although these are important measures for reading performance, there are many text characteristics that are known to differ by genre (e.g., pronoun use, van Hell et al., 2005; connectives; van Silfhout et al., 2015; vocabulary; Gardner, 2004) that may be important in inferential comprehension.

Standards of coherence become more sophisticated with age (van den Broek & Kendeou, 2017); therefore, the age of the reader (child or adult) was considered as a potential moderator. However, the analyses did not indicate that the age of the reader moderated the effect of genre on

inferential comprehension. This contrasts with previous work finding that listening comprehension for narratives improved with age, but there were no differences in expository texts (Lehto & Anttila, 2003). However, there has been limited examination of inferential reading comprehension differences with age and the work that has been done has been with narrative texts (e.g., Franks, 1997; Kendeou et al., 2008; van den Broek et al., 2001). It may be that inferential reading comprehension develops similarly for narrative and expository texts as readers mature. However, as with inference type and text readability issues, it is unclear if the lack of a finding in this moderator analysis is due to lack of power rather than an actual lack of an effect (Hempel et al., 2013). Potentially, adults may have more sophisticated standards of coherence regarding genre than children do and as such may have different inferential comprehension.

Limitations and Future Directions

One key limitation of this meta-analysis is that it is unclear what aspects of genre may explain why this difference in inferential comprehension was noted. Although there was an insufficient number of studies for a moderator analysis, one potential reason could be the structure of texts within the genres. It is possible that readers would have similar levels of inferential comprehension for narrative and certain subtypes of expository texts that have clearer structures or structures with similarities to narratives, such as biographies. Another possibility is that the variations in expository text structures is additional information for readers to process, thereby limiting their cognitive resources that could be applied to inferential comprehension (Hebert et al., 2016). Instruction in expository text structures, which has been shown to improve expository text comprehension (Hebert et al., 2016; Pyle et al., 2017), may ameliorate this issue and could reduce the genre difference in inferential comprehension noted in this meta-analysis.

Moreover, there is a need for more inquiry into improving inferential comprehension of expository texts. In a meta-analysis of 33 studies on inference instruction, only 30% of the studies reported using expository texts (Elleman, 2017).

Another potential reason for the genre difference in inferential comprehension is related to the linguistic features of the texts. Some of the studies reported matching the texts for readability based on linguistic features and whether or not the texts were matched was not identified as a moderator. However, the readability measures were generally limited to examining word and sentence length (see Table 1). A more fine-grained approach that examines linguistic features such as pronouns, connectives, and vocabulary known to be important for text comprehension may be informative. Such an approach could potentially identify linguistic features that are mechanisms of the genre differences in inferential comprehension. If such features could be identified, this would inform the design and writing of expository texts to facilitate inferential comprehension. In a similar approach, current guidelines for simplifying expository texts could be tested for their effects on inferential comprehension (Arfé et al., 2018).

There are additional limitations of this meta-analysis that should be noted. Only reports written in English were examined; therefore, there could be relevant findings disseminated in other languages that were not included. Several reports did not indicate the reliability of the inferential measures reported, which makes it subsequently difficult to determine the validity of the effect sizes. Because the texts in the genre comparisons were almost always different (Zwaan, 1994, being an exception), the inferential comprehension measures were frequently different. In this way, it cannot be determined if the inferential comprehension effects noted were due to the texts or the measures. Finally, the variability in inferential comprehension by genre could be related to differences in reading skill (e.g., Karlsson et al., 2018; Kraal et al., 2019).

However, there was insufficient information in the reports regarding reading skill levels to test this possibility with a moderator analysis.

Conclusion

In order to comprehend a text, the ideas must be connected through inferences. This process has been assumed to be more difficult with expository texts than narrative texts; the findings from this meta-analysis support this assumption. However, based on the moderator analyses, it is unclear why this genre difference is the case. Future work may address this by examining how variations in expository text structure, genre differences in linguistic features, or inference instruction focused on expository texts may address the genre difference in inferential comprehension found in this meta-analysis.

References

*indicates inclusion in meta-analysis

Arfé, B., Mason, L., & Fajardo, I. (2018). Simplifying informational text structure for struggling readers. *Reading and Writing: An Interdisciplinary Journal*, *31,* 2191-2210. https://doi.org/10.1007/s11145-017-9785-6

Baker, L., & Stein, N. (1981). The development of prose comprehension skills. In C. Santa, B. Hayes (Eds.), *Children's prose comprehension: Research and practice*, International Reading Association.

Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing: An Interdisciplinary Journal*, *24,* 183–202. https://doi.org/10.1007/s11145-010-9264-9

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*, *29,* 137–164. https://doi.org/10.1080/02702710801963951

*Best, R., Ozuru, Y., Floyd, R. G., & McNamara, D. (2006). Children's text comprehension: Effects of genre, knowledge, and text cohesion. In *ICLS 2006 - International Conference of the Learning Sciences, Proceedings* (Vol. 1, pp. 37-42)

Bloch, M. (2014). Meta-analysis and moderator analysis: Can the field develop further? *Journal of The American Academy of Child & Adolescent Psychiatry, 53,* 135-137. https://doi.org/10.1016/j.jaac.2013.12.001

Bohn-Gettler, C. M., & Kendeou, P. (2014). The interplay of reader goals, working memory, and text structure during reading. *Contemporary Educational Psychology*, *39,* 206–219. https://doi.org/10.1016/j.cedpsych.2014.05.003

Borenstein, M., Hedges, L. V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to meta-analysis*. Wiley.

*Bowyer-Crane, Claudine (2002) *The relationship between reading comprehension and online inference generation in children* (Doctoral thesis). Retrieved from http://etheses.whiterose.ac.uk/id/eprint/14165

*Bridge, C. A., & Tierney, R. J. (1981). The inferential operations of children across text with narrative and expository tendencies. *Journal of Reading Behavior*, *13,* 201–214. https://doi.org/10.1080/10862968109547408

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, *11*(5-6), 489-503. https://doi.org/10.1023/A:1008084120205

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96,* 31. https://doi.org/10.1037/0022-0663.96.1.31

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96,* 671. https://doi.org/10.1037/0022-0663.96.4.671

Carlisle, J. F. (1989). The use of the sentence verification technique in diagnostic assessment of listening and reading comprehension. *Learning Disability Research, 5,* 33–44.

Clinton, V., & Van den Broek, P. (2012). Interest, inferences, and learning from texts. *Learning and Individual Differences*, *22,* 650–663. https://doi.org/10.1016/j.lindif.2012.07.004

Collins, A. A., Compton, D. L., Lindström, E. R., & Gilbert, J. K. (in-press). Performance

    variations across reading comprehension assessments: Examining the unique

    contributions of text, activity, and reader. *Reading and Writing: An Interdisciplinary

    Journal,* https://doi.org/10.1007/s11145-019-09972-5

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency:

    The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35,*

    115–135. https://doi.org10.1111/j.1467-9817.2010.01449.x

*De Beni, R., Palladino, P., Borella, E., Presti, S. (2003). Reading comprehension and aging:

    Does an age-related difference necessarily mean impairment? *Aging Clinical and

    Experimental Research*, *15,* 67–76. https://doi.org/10.1007/BF03324482

*De Beni, R., Borella, E., & Carretti, B. (2007). Reading comprehension in aging: The role of

    working memory and metacomprehension. *Aging, Neuropsychology, and Cognition*, *14,*

    189–212. https://doi.org/1080/13825580500229213

*Denton, C. A., Enos, M., York, M. J., Francis, D. J., Barnes, M. A., Kulesz, P. A., Fletcher, J.

    M., & Carter, S. (2015). Text-processing differences in adolescent adequate and poor

    comprehenders reading accessible and challenging narrative and informational

    text. *Reading Research Quarterly*, *50,* 393–416. https://doi.org/10.1002/rrq.105

Elleman, A. M. (2017). Examining the impact of inference instruction on the literal and

    inferential comprehension of skilled and less skilled readers: A meta-analytic

    review. *Journal of Educational Psychology*, *109,* 761–781.

    https://doi.org/10.1037/edu0000180

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting

    think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and*

    Activity, *5,* 178–186. https://doi.org/10.1207/s15327884mca0503_3

Fisher, Z., & Tipton, E. (2014). robumeta: An R-package for robust variance estimation in meta-

    analysis.

Follmer, D. J. (2018). Executive function and reading comprehension: A meta-analytic

    review. *Educational Psychologist*, *53,* 42–60.

    https://doi.org/10.1080/00461520.2017.1309295

Franks, B. A. (1997). Deductive reasoning with prose passages: Effects of age, inference form,

    prior knowledge, and reading skill. *International Journal of Behavioral Development*, *21,*

    501–535.

Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found

    in children's narrative and expository reading materials. *Applied Linguistics, 25,* 1–37.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32,* 99–

    156. https://doi.org/10.1016/0010-0277(89)90001-2

Graesser, A. C., Hauft-Smith, K., Cohen, A. D., & Pyles, L. D. (1980). Advanced outlines,

    familiarity, and text genre on retention of prose. *The Journal of Experimental*

    Education, *48,* 281–290. https://doi.org/10.1080/00220973.1980.11011745

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing

    multilevel analyses of text characteristics. *Educational Researcher*, *40,* 223–234.

    https://doi.org/10.3102/0013189X11413260

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,* 193–202. https://doi.org/10.3758/BF03195564

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text

  comprehension. *Psychological Review*, *101,* 371–395.

  https://doi.org/10.1037/0033-295X.101.3.371

Griffiths, G. G., Sohlberg, M. M., Kirk, C., Fickas, S., & Biancarosa, G. (2016). Evaluation of

  use of reading comprehension strategies to improve reading comprehension of adult

  college students with acquired brain injury. *Neuropsychological Rehabilitation*, *26,*

  161-190. https://doi.org/10.1080/09602011.2015.1007878

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and

  some of their interactions. *Journal of Experimental Psychology: General*, *114,* 357–

  374. https://doi.org/10.1037/0096-3445.114.3.357

Hamilton, S., Freed, E., & Long, D. L. (2016). Word-decoding skill interacts with working

  memory capacity to influence inference generation during reading. *Reading Research*

  Quarterly, *51,* 391-402. https://doi.org/10.1002/rrq.148

Hebert, M., Bohaty, J. J., Nelson, J. R., & Brown, J. (2016). The effects of text structure

  instruction on expository reading comprehension: A meta-analysis. *Journal of*

  Educational Psychology, *108,* 609. https://doi.org/10.1037/edu0000082

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related

  estimators. *Journal of Educational Statistics, 6,* 107–128.

  https://doi.org/10.3102/10769986006002107

Hempel, S., Miles, J. N., Booth, M. J., Wang, Z., Morton, S. C., & Shekelle, P. G. (2013). Risk

    of bias: A simulation study of power to detect study-level moderator effects in meta-

    analysis. *Systematic Reviews, 2,* 107. https://doi.org/10.1186/2046-4053-2-107

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of*

    interventions. Version 5.1.0*.* The Cochrane Collaboration. www.handbook.cochrane.org.

*Hynd, C. R., & Chase, N. D. (1991). The relation between text type, tone, and written

    response. *Journal of Reading Behavior*, *23,* 281–306.

    https://doi.org/10.1080/10862969109547743

*Karlsson, J., van den Broek, P., Helder, A., Hickendorff, M., Koornneef, A., & van

    Leijenhorst, L. (2018). Profiles of young readers: Evidence from thinking aloud while

    reading narrative and expository texts. *Learning and Individual Differences*, *67,* 105–116.

    https://doi.org/10.1016/j.lindif.2018.08.001

Kendeou, P. (2015). A general inference skill. In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr.

    (Eds.), *Inferences during reading.* (pp. 160–181). Cambridge University Press.

Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference

    generation across different media. *Journal of Research in Reading*, *31,* 259–272.

    https://doi.org/10.1111/j.1467-9817.2008.00370.x

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

*Kraal, A., Koornneef, A. W., Saab, N., & van den Broek, P. W. (2018). Processing of

    expository and narrative texts by low- and high-comprehending children. *Reading*

    and Writing: An Interdisciplinary Journal, *31,* 2017–2040.

    https://doi.org/10.1007/s11145-017-9789-2

Kraal, A., van den Broek, P. W., Koornneef, A. W., Ganushchak, L. Y., & Saab, N. (2019). Differences in text processing by low-and high-comprehending beginning readers of expository and narrative texts: Evidence from eye movements. *Learning and Individual Differences*, *74,* 101752. https://doi.org/10.1016/j.lindif.2019.101752

Kulesz, P. A., Francis, D. J., Barnes, M. A., & Fletcher, J. M. (2016). The influence of properties of the test and their interactions with reader characteristics on reading comprehension: An explanatory item response study. *Journal of Educational Psychology*, *108,* 1078. https://doi.org/10.1037/edu0000126

*Lee, H. C. (2014). Inferencing behaviour of ESL readers. *Electronic Journal of Foreign Language Teaching*, *11,* 21-37.

Lehto, J. E., & Anttila, M. (2003). Listening comprehension in primary level grades two, four and six. *Scandinavian Journal of Educational research*, *47,* 133–143. https://dx.doi.org/10.1080/00313830308615

Lorch, R. F. (2015). What about expository text? In E. J. O'Brien, A. E. Cook, & R. F. Lorch (Eds.). Inferences during reading (pp. 348–361). Cambridge University Press.

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91,* 615. https://doi.org/10.1037/0022- 0663.91.4.615

*Margolin, S. J., Driscoll, C., Toland, M. J., & Kegler, J. L. (2013). E-readers, computer screens, or paper: Does reading comprehension change across media platforms? *Applied Cognitive Psychology*, *27,* 512-519. https://doi.org/10.1002/acp.2930

McMaster, K. L., Van den Broek, P., Espin, C. A., White, M. J., Rapp, D. N., Kendeou, P., Bohn-Gettler, C.M., & Carlson, S. (2012). Making the right connections: Differential

effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, *22,* 100-111. https://doi.org/10.1016/j.lindif.2011.11.017

McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Sources of text difficulty: Across genres and grades. In J. P. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 89–116). R&L Education.

Meyer, B. J. F. (1987). Following the author's top-level organization: An important skill for reading comprehension. In R. J.Tierney, P. L.Anders, & J.Nichols Mitchell (Eds.), *Understanding readers' understanding: Theory and practice* (pp. 59–76). Erlbaum.

*Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology*, *91,* 488–496. https://doi.org/10.1037/0022-0663.91.3.488

*Olson, M. W. (1985). Text type and reader ability: The effects on paraphrase and text-based inference questions. *Journal of Reading Behavior*, *17,* 199–214. https://doi.org/10.1080/10862968509547540

Oudega, M., & van den Broek, P. (2018). Standards of coherence in reading variations in processing and comprehension of text. In Millis, K., Long, D., Magliano, J., Wiemer, K. (Eds.), *Deep Comprehension: Multidiciplinary Approaches to Understanding, Enhancing, and Measuring Comprehension.* (pp. 41-51). Routledge.

Perfetti, C. A., & Stafura, J. Z. (2015). Comprehending implicit meanings in text without making inferences. In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading.* (pp. 1–18). Cambridge University Press.

Polanin, J. R., Hennessy, E. A., & Tanner-Smith, E. E. (2017). A review of meta-analysis

packages in R. *Journal of Educational and Behavioral Statistics*, *42,* 206-242.

https://doi.org/10.3102/1076998616674315

Pressley, M., & Afflerbach, P. (1995). Verbal reports of reading: The nature of constructively

responsive reading. Erlbaum.

Pyle, N., Vasquez, A. C., Lignugaris, K. B., Gillam, S. L., Reutzel, D. R., Olszewski, A., Segura,

H., Hartzheim, D., Liang, W. & Pyle, D. (2017). Effects of expository text structure

interventions on comprehension: A   meta-analysis. *Reading Research Quarterly*, *52,*

469–501. https://doi.org/10.1002/rrq.179

*Saadatnia, M., Ketabi, S., & Tavakoli, M. (2017). Levels of reading comprehension across

text types: A comparison of literal and inferential comprehension of expository and

narrative texts in Iranian EFL learners. *Journal of Psycholinguistic Research*, *46,* 1087–

1099. https://doi.org/10.1007/s10936-017-9481-3

*Sáenz, L. M., & Fuchs, L. S. (2002). Examining the reading difficulty of secondary students

with learning disabilities: Expository versus narrative text. *Remedial and Special

Education*, *23,* 31–41. https://doi.org/10.1177/074193250202300105

Seipel, B., Carlson, S. E., & Clinton, V. E. (2017). When do comprehender groups differ? A

moment-by-moment analysis of think-aloud protocols of good and poor

comprehenders. *Reading Psychology*, *38,* 39–70.

https://doi.org/10.1080/02702711.2016.1216489

Singer, M. (2007). Inference processing in discourse comprehension. In G. Gaskell (Eds.),

Oxford handbook of psycholinguistics (pp. 343-359). Oxford University Press.

Singer, M., Andruslak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging

    inference processes. *Memory & Cognition*, *20,* 539–548.

    https://doi.org/10.3758/BF03199586

Schmitz, A., Gräsel, C., & Rothstein, B. (2017). Students' genre expectations and the effects of

    text cohesion on reading comprehension. *Reading and Writing: An Interdisciplinary*

    Journal, *30,* 1115–1135. https://doi.org/10.1007/s11145-016-9714-0

*Strickland, B., Barrie, S., Williams, R. (2011). Discourse structure and word learning.

    Pragmatics and Society, *2,* 260–281. https://doi.org/10.1075/ps.2.2.07str

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data

    structures using robust variance estimates: A tutorial in R. *Journal of Developmental and*

    Life-Course Criminology, *2,* 85-112. https://doi.org/10.1007/s40865-016-0026-5

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and

    model fit using robust variance estimation in meta-regression. *Journal of Educational*

    and Behavioral Statistics, *40,* 604–634. https://doi.org/10.3102/1076998615606099

van den Broek, P., Bohn-Gettler, C. M., Kendeou, P., Carlson, S., & White, M. J. (2011). When

    a reader meets a text: The role of standards of coherence in reading comprehension. In M.

    T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from*

    text.(pp. 123–139). IAP Information Age Publishing.

van den Broek, P., & Gustafson, M. (1999). Comprehension and memory for texts: Three

    generations of reading research. In S. R. Goldman, A. C. Graesser, & P. van den Broek

    (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom*

    Trabasso. (pp. 15–34). Lawrence Erlbaum Associates Publishers.

van den Broek, P., Kendeou, P., Lousberg, S., & Visser, G. (2017). Preparing for reading

comprehension: Fostering text comprehension skills in preschool and early elementary

school children. *International Electronic Journal Of Elementary Education, 4,* 259-

268. https://www.iejee.com/index.php/IEJEE/article/view/223

van den Broek, P., Tzeng, Y., Risden, K., Trabasso, T., & Basche, P. (2001). Inferential

questioning: Effects on comprehension of narrative texts as a function of grade and

timing. *Journal of Educational Psychology, 93,* 521-529.

https://doi.org/10.1037/0022-0663.93.3.521

van den Broek, P., White, M. J., Kendeou, P., & Carlson, S. (2009). Reading between the lines:

Developmental and individual differences in cognitive processes in reading

comprehension. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.),

Beyond decoding: The behavioral and biological foundations of reading comprehension

(pp. 107-123). Guilford Press.

van Hell, J. G., Verhoeven, L., Tak, M., & Van Oosterhout, M. (2005). To take a stance: a

developmental study of the use of pronouns and passives in spoken and written narrative

and expository texts in Dutch. *Journal of Pragmatics*, *37,* 239-273.

https://doi.org/10.1016/S0378-2166(04)00190-0

van Silfhout, G., Evers-Vermeul, J., & Sanders, T. (2015). Connectives as processing signals:

How students benefit in processing narrative and expository texts. *Discourse

Processes*, *52,* 47-76. https://doi.org/10.1080/0163853X.2014.905237

Virtue, S., van den Broek, P., & Linderholm, T. (2006). Hemispheric processing of inferences:

The effects of textual constraint and working memory capacity. *Memory &

Cognition*, *34,* 1341-1354. https://doi.org/10.3758/BF03193276

Walkington, C., Clinton, V., Ritter, S. N., & Nathan, M. J. (2015). How readability and topic

incidence relate to performance on mathematics story problems in computer-based

curricula. *Journal of Educational Psychology*, *107,* 1051–1074.

https://doi.org/1037/edu0000036

Walkington, C., Clinton, V., & Sparks, A. (2019). The effect of language modification of

mathematics story problems on problem solving in online homework. *Instructional

Science, 47,* 499–529. https://doi.org/10.1007/s11251-019-09481-6

Williams, J. P. (1993). Comprehension of students with and without learning disabilities:

Identification of narrative themes and idiosyncratic text representations. *Journal of

Educational Psychology*, *85,* 631–641. https://doi.org/10.1037/0022-0663.85.4.631

*Wolfe, M. B. W., & Woodwyk, J. M. (2010). Processing and memory of information presented

in narrative or expository texts. *British Journal of Educational Psychology*, *80,* 341–

362. https://doi.org/10.1348/000709910X485700

*Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of

Experimental Psychology: Learning, Memory, and Cognition*, *20,* 920–933.

https://doi.org/10.1037/0278-7393.20.4.920

Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. *Handbook of Psycholinguistics*,

2, 725-764. https://doi.org/10.1016/B978-012369374-7/50019-5

Table 1

Description of Studies in the Meta-Analysis

Author/s (year)	Number of Participants (age level)	Texts	Text Matching	Types of Inferences: Inferential measure	Reliability of Inferential Measure	Findings
Best et al. (2006)	64 (children; 4th grade)	2 narrative ("*Moving*"; "*Orlando*") and 2 expository ("*Heat*"; "*Needs of Plants*") texts (401 to 634 words)	No	Text-connecting: Multiple choice questions after reading	Not described	Children's comprehension was better for narrative texts, as compared to expository texts
Bridge & Tierney (1981)	36 (children, 3rd grade)	1 Narrative ("*Johnny and the Squirrel*" 153 words) and 1 expository ("*Dinosaurs*") text (190 words)	No	Text-connecting: Free and probed recall after reading	Not described	Both good and poor readers were able to cohesively recall more information for the narrative text as compared to the expository text
Bowyer-Crane (2002) Exp 2	22 (children, 5th grade)	32 narrative and 32 expository (each 6 sentences long)	Yes: number of words; same target inference sentences	Causal bridging and elaborative (grouped together in analyses): Sentence verification after reading	Not described	More correct responses were given to narrative text sentences than expository text sentences

				Knowledge-based: Comprehension questions after sentence verification		
De Beni et al. (2003)	250 (adults; 55 to 90 years old)	1 Narrative ("*Deafness*" 500 words) and 1 expository ("*Describing an experiment in Japan to increase understanding of family dependence on mass media*") text (400 words)	No	Text-connecting: Multiple-choice questions after reading	Test-retest and internal consistency based on standardized reading assessment. Details not stated.	More correct responses to inferential questions about the narrative than the expository texts for both young-old and old-old age groups
De Beni (2007)	90 (adults; 20 to 85 years old)	1 Narrative ("*Deafness*" 40 lines) and 1 expository ("*Japan: Leaders in the cartoon world*") text (25 lines)	No	Text-connecting: Multiple-choice questions after reading	Same materials as the 2003 article by the same first author	More correct responses to inferential questions about the narrative than the expository texts for both young-old and old-old age groups
Denton et al. (2016)	325 (adolescents; 12 to 18 years old)	45 narrative and 44 expository	No	Text-connecting and knowledge-based (grouped together in	Inter-rater reliability was high	More inferences for narrative than expository, but only for accessible

				analyses): Think aloud while reading		texts. Inferences were similar by genre for challenging texts.
Hynd & Chase (1991)	58 (adults; college freshman)	1 narrative, 1 expository (both on suburban development, each with personal and impersonal versions)	No	Explanatory: Essay response after reading	Inter-rater reliability was high	Similar findings by genre for inference
Karlsson et al. (2018)	107 (children; 9-11 years old)	2 narrative (16-17 sentences each) and 2 expository texts (15-16 sentences each)	No	Text-connecting, elaborative, and predictive: Think aloud during reading	Inter-rater reliability was high	Overall, more inferences while reading narrative texts than expository. Reading profiles (literal, paraphrasing, and elaborating readers) tend to be similar across genre
Kraal et al. (2018)	87 (children, 2nd grade)	2 Narrative and 2 expository texts (average 146 words each)	Yes, text difficulty was the same grade level (based on word and sentence length, percentage of high frequency	Valid inference (elaborative and predictive grouped together): Think aloud during reading Gap-filling and text-connecting: Comprehension	Inter-rater reliability was high	More inferences for narrative than expository texts. More accurate answers for inferential questions after reading narrative than expository.

			words, and lexical diversity) and the text length was the same	questions after reading		
Lee (2014)	9 (children; 6th grade)	1 Narrative ("*The travellers and the bear*") and 1 expository ("*Green buildings*") (~80 words each)	No	Bridging, text-connecting, explanatory, and elaborative: Think aloud during reading	Inter-rater reliability was high	Fewer inferences overall for expository than narrative text.
Margolin et al. (2013)	90 (adults; 18 to 25 years old)	5 Narrative (from literacy books in library) and 5 expository (biographies of historical pop culture figures) texts (492 to 572 words)	Yes, similar Flesch-Kincaid grade levels	Inference type not specified: Multiple choice questions after reading	Good internal consistency	Inferential comprehension scores for narrative passages were lower than for expository passages
Narvaez et al. (1999)	20 (adults; college students)	2 Narrative ("*Spa*"; "*All about Suicide*" 34 and 31 sentences) and 2 expository ("*Comet Fire*"; "*Viruses*") (19	No	Explanatory and predictive: Think aloud during reading	Inter-rater reliability was high	There were more explanations and predictions for the narrative than the expository text.

		and 23 sentences)				
Olson (1985)	53 (children; 3rd grade)	2 Narrative ("*A Winter Bouquet*"; "*Here Comes the Landlord!*") and 2 expository ("*Beware the Plant*"; "*Where Does the Rain Come From?*") texts (334-353 words)	Yes, similar average word frequency	Text-based: Comprehension questions after reading	Inter-rater reliability was high	Inferential questions were more difficult to answer for the expository texts than the narrative.
Saadatnia et al. (2017)	180 (adults; college students)	4 narrative and 4 expository texts (details not available)	No	Inference type not specified: Multiple choice questions after reading	Not described	There was no significant difference in inferential comprehension between the two text types.
Saenz & Fuchs (2002)	111 (children; high school students)	4 narrative (345-427 words) and 4 expository (378-434 words)	Yes, similar Flesch-Kincaid readability grade levels and passage lengths	Inference type not specified: Comprehension questions after reading	Inter-rater reliability was high	A significant effect was found between text types for inferential questions with narrative inferential comprehension being superior to expository

						inferential comprehension.
Strickland (2011)	74 (adults; college students)	6 narrative (average 109 words each) and 6 expository (average 98 words each)	Yes, similar Lexile ratings (based on word frequency and average sentence length) by genre	Incidental vocabulary acquisition (inferring word meaning from context): Definition production test after reading	Not described	Better inferring of word meanings from context in narrative compared to expository texts.
Wolfe & Woodwyk (2010) Exp 1	61 (adults; college students)	1 Narrative ("*Alex's Adventure*" 444 words) and 1 expository (Topic of the human circulatory system) (411 words)	Yes, both on the same topic of human circulation with matched common sentences. Similar causal connections between genre, although .	Text-connecting and knowledge-based: Think aloud during reading	Inter-rater reliability was high	More knowledge-based inferences for the expository text than the narrative for both common and non-common sentences. No differences in text-connecting inferences.
Zwaan (1994), Exp 1	38 (adults; college students)	6 texts, 4 excerpts from news articles and two excerpts from novels (average 216 words each)	Yes, same texts with participants assigned to different genre expectations (news articles or literary fiction)	Inference type not specified: Sentence verification after reading	Not described	Situation model construction was weaker when readers thought texts were literary fiction compared to readers who thought texts were news articles

| Zwaan (1994) Exp 2 | 40 (adults; college students) | 4 of the texts used in Zwaan (1994)'s Exp 1 | Yes, same texts with participants assigned to different genre expectations (news articles or literary fiction) | Inference type not specified (separate measures for causally important and nonimportant): Sentence verification after reading | Not described | Similar to Experiment 1, situation model construction was weaker when readers thought texts were literary fiction compared to readers who thought texts were news articles |

Table 2

Inferential Comprehension Statistics for Each Study and Model Statistics (positive Hedges' g indicates a higher score for the narrative texts compared to the expository texts)

Study Name	Hedges' *g*	variance	Sample Size		
			N	E	T
Best et al. (2006) Global questions	1.33	.03			64
Bridge & Tierney (1981)					36
Free	.90	.02			
Probed	-.10	.01			
Bowyer-Crane (2002) Exp 2					
Target inference sentences, first set true	.28 .41	.02 .02			
Target inference sentences, first set false	-.10	.02			
Target inference sentences, second set true	.23	.02			
Target inference sentences, second set false	-.40	02			22
Comprehension questions					
De Beni et al. (2003) Multiple choice questions	.81	.003			250
De Beni et al. (2007) Multiple choice questions	-.68	.01			90
Denton et al. (2016)					
Accessible text	.47	.01			
Challenging text	.78	.01			325
Hynd & Chase (	.37	.01			58
Karlsson et al.					
Text-connecting inference	.27	.00			
Valid elaborative inference	-.01	.00			
Predictive inference	.52	.00			107
Kraal et al.					
Valid inference	.43	.01			
Text-connecting question	.52	.01			
Gap-filling question	.49	.01			87
Lee					
Bridging inference	12.11	3.30			
Global inference	-2.65	.19			
Explanation	8.74	.173			
Elaboration	.00	.04			9
Margolin et al.	-.25	.00			90
Narvaez et al.					

Explanations	.99	.03			
Predictions	.51	.02			20
Olson					
Information	1.75	.02			
Logical	1.21	.01			53
Saadatnia	.07	.00			180
Multiple choice questions					
Saenz & Fuchs	.53	.00			111
Inferential questions					
Strickland et al.	.26	.01			74
Incidental word acquisition					
Wolfe & Woodwyk					
Prior text elaboration, common sentences	.08	.07			
Prior knowledge elaboration, common sentences	.59	.07			
Prior text elaboration, non-common sentences	.08	.07			
Prior knowledge elaboration, non-common sentences	.78	.07	30	30	60
Zwaan (1994) Exp 1	-.69	.12	19	19	38
Sentence verification					
Zwaan (1994) Exp 2					
Sentence verification, important	-.55	.10			
Sentence verification, unimportant	-.36	.10	20	20	40

Note. Most studies were within subjects and all participants read narrative and expository texts.

Table 3

Sensitivity Analysis of Findings with Varying Correlations Assumed for Dependent Effect Sizes

Rho	0	.2	.4	.6	.8	1
Hedges' g	.36	.36	.36	.36	.36	.36
Standard Error	.14	.14	.14	.14	.14	.14

Table 4

Meta-regression results

	Beta	SE	T	Dfs	p	95% CI Lower	95% CI Upper
Intercept	.45	.16	2.82	7.9	.02	.08	.82
Timing	-.21	.28	-.77	8.02	.46	-.85	.43
Matched	-.14	.32	-.45	11.65	.66	-.83	.55
Age	-.33	.25	-1.30	12.58	.22	-.87	.22

	Beta	SE	T	Dfs	p	95% CI Lower	95% CI Upper
Intercept	.70	.34	2.09	6.32	.08	-.11	1.52
Type	-.41	.43	-.96	7.43	.36	-1.41	.59

Note. Timing = inferential comprehension measure was during reading (coded 0) or after reading (coded 1). Matched = texts of different genres were matched for readability (coded 0 for no, 1 for yes). Age = readers were children or adults (coded 0 for children, 1 for adults). Type = inferential measure was text connecting (coded = 0) or knowledge based (coded = 1). SE = standard error. T = t-test value. Dfs = degrees of freedom. 95% CI Lower = 95% confidence interval lower limit. 95% CI Upper = 95% confidence interval upper limit. For the first model, k = 19, for the second model k = 11.
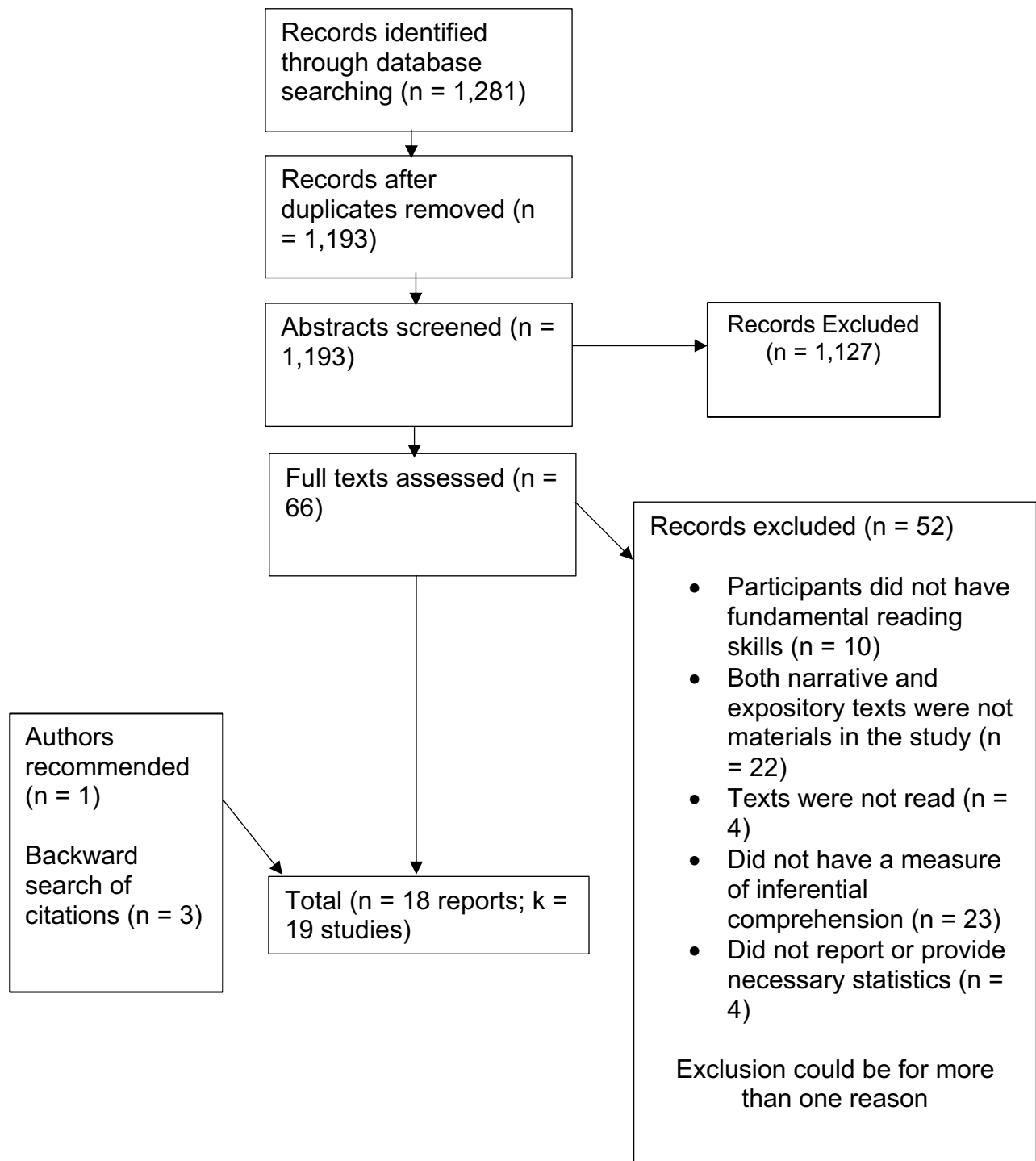
Records identified through database searching (n = 1,281)

Records after duplicates removed (n = 1,193)

Abstracts screened (n = 1,193)

Records Excluded (n = 1,127)

Full texts assessed (n = 66)

Records excluded (n = 52)

- Participants did not have fundamental reading skills (n = 10)
- Both narrative and expository texts were not materials in the study (n = 22)
- Texts were not read (n = 4)
- Did not have a measure of inferential comprehension (n = 23)
- Did not report or provide necessary statistics (n = 4)

Exclusion could be for more than one reason

Authors recommended (n = 1)

Backward search of citations (n = 3)

Total (n = 18 reports; k = 19 studies)

Figure 1. Flow diagram of the systematic review process