The Effect Logical Relatedness and Semantic Overlap on Argument Evaluation

Srikanth Dandotkar

Eastern Illinois University

Joseph P. Magliano and M. Anne Britt

Northern Illinois University

Accepted for Publication in Discourse Processes

Author Note

Srikanth Dandotkar, Department of Psychology, Eastern Illinois University.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education in part by Grant R305F100007 Reading for Understanding Across Grades 6 through 12: Evidence-based Argumentation for Disciplinary Learning from the Reading for Understanding Research Initiative to M. Anne Britt and Joseph P. Magliano. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

Correspondence concerning this article should be addressed to Srikanth Dandotkar, Department of Psychology, Eastern Illinois University, Charleston IL 61920.

Contact: sdandotkar@eiu.edu

Abstract

In two studies, we examined the extent to which skilled and less-skilled reasoners of arguments relied on relevance relations (semantic & logical relatedness) between claims and reasons when evaluating arguments. College students, selected as having high or low analytical reasoning skill, evaluated the quality of a set of two sentence arguments and rated the strength of their agreement with them. The arguments were structured to vary in their degree of semantic and logical relatedness. Experiment 1 used a direct test of readers' evaluation of the logical connection while Experiment 2 used a more naturalistic, multifactor task (agreement judgment). Overall, both skilled and less-skilled reasoners accepted and more strongly agreed with high logical arguments than they did with low logical arguments. This indicates that both skilled and less-skilled reasoners rely more on logical relatedness when evaluating arguments. However, skilled reasoners' reliance on logical relatedness was higher than that of less-skilled reasoners,' particularly when evaluating the quality of and not their strength of agreement with arguments. With respect to semantic relatedness, readers' reliance on this factor was minimal across experiments. Contrary to the findings in the narrative comprehension research, less-skilled reasoners did not excessively rely on semantics under low logical conditions. In fact, skilled evaluators relied on semantics more when evaluating the quality of high logical arguments. Finally, the current studies confirm the earlier findings that precision in remembering the main verb of the claim is fundamental to one's skill in evaluating arguments.

The Effect of Logical Relatedness and Semantic Overlap on Argument Evaluation

Being able to comprehend, evaluate, and produce arguments is a critical aspect of academic literacy (Kuhn, 1993; Larson, Britt, & Kurby, 2009; Perkins, Faraday, Bushey, 1991; Voss, Wiley, & Carretero, 1995). Argumentation skills are particularly important for successful performance in college. College students require these skills to write research papers, comprehend information from different sources, evaluate models or theories, engage in class discussions and answer essay questions on exams (Wolfe, 2011). For this reason, we expect students to learn these skills in high school (Achieve, Inc., 2013; Council of Chief State School Officers, 2010). Yet, many children who graduate from high school still struggle to comprehend, evaluate and produce arguments (National Science Board, 2006; National Assessment of Educational Progress, 2011; National Assessment of Educational Progress, 2013). Both high school and college students have difficulty in evaluating the quality of simple arguments (Larson et al., 2009). For example, Larson et al. (2009) had participants evaluate simple arguments like 1a (reason provides support) and 1b (reason does not provide support) for whether or not the arguments were minimally structurally acceptable. They found that college students' performance (60-65% accuracy), while better than chance, was far from mastery. For educators to help them, it is important that we understand which factors successful students rely on when evaluating arguments such as ones presented below.

1a. Vaccinations should be required for all children because it is important to protect kids from infectious diseases.

1b. Vaccinations should be required for all children because inoculations contain weakened viruses.

1c. Vaccinations should be required for all children because it is important to protect humans from serious problems.

1d. Vaccinations should be required for all children because of the presence of weakened organisms.

Two potential sources that readers can attend to for evaluating arguments are logical connection (can the reason support the claim) and lower level semantic overlap (lexical or word-to-word relatedness). To illustrate, the claims for the arguments in 1a-1d are the same ("Vaccination should be required for all children") but vary in how the reasons are related to the claim. The reasons in 1a and 1c are each logically related to the claim in that things that provide protection from harm can justify requiring people to take an action (high logical relatedness). In contrast, the reasons provided in arguments 1b and 1d, while factually true, are not relevant to whether we ought to require people to be vaccinated (low logical relatedness). Reasons may also vary in the degree of semantic overlap at the lexical level. Arguments 1a and 1b have high lexical overlap, several words (for example, "kids" and "infectious diseases" or "inoculations" and "weakened viruses") in the reason are lexically related to words in the claim (for example, "vaccinations," "children"). Arguments 1c and 1d, in contrast, have lower lexical overlap.

In two studies, we examined whether logical relatedness and lexical overlap between claims and reasons of arguments affect how skilled and less-skilled reasoners evaluate arguments. Obviously, one would wish students to evaluate the quality of arguments based on the strength of the logical relationships that bind claims and reasons. However, to the extent that readers use lexical overlap in evaluating arguments, they may be duped into accepting poor arguments.

Logical Relatedness

Arguments can be evaluated in different ways. One might assess one's agreement with an argument as a whole (Britt, Kurby, Dandotkar, & Wolfe, 2008; Voss, Fincher-Kiefer, Wiley & Silfies, 1993; Wolfe & Britt, 2008; Wolfe et al., 2009), based on a general sense that the argument is "good" and coherent with one's beliefs. Alternatively, one might evaluate the quality of support (Kuhn, 1993; Larson et al., 2009; Shaw, 1996; Wolfe, Britt, & Butler, 2009), assessing whether provided reasons are "true" (i.e., criteria of acceptability), whether the structural relationship between a reason and its claim is logically relevant (i.e., criteria of relevance), and whether the reasons constitute sufficient evidence (i.e., criteria of sufficiency) (Johnson & Blair, 1977). Here, we focus on readers' ability to evaluate the structural relationship of arguments because this is the essence of an argument (Angell, 1964; Johnson & Blair, 1977; Freeman, 1991; Salmon, 1984; Toulmin, 1958; Voss & Means, 1991).

People can assess the quality of an argument as they read, but this depends on how carefully they read and represent the information. Britt and Larson (2003), for example, found that when readers detect that they are reading an argument, they can use the claim to organize supporting and opposing information. Once represented, potential reasons can then be evaluated against the claim to determine whether or not they are relevant. However, for this evaluation to be accurate, the reader must build a precise representation of the claim predicate (Britt et al., 2008). For instance, after reading the claim in arguments 1a-d, a precise representation would be that the author is asserting "should be required" as the predicate. A reader with only a gist representation for the claim (for example, "is good" or "is needed") might regard 1c as a perfectly acceptable argument. Britt et al. (2008) found that readers who have a less precise memory for the main predicate of the claim are less accurate in evaluating the logical structure of

arguments, that is, correctly determining whether a reason is relevant to a claim. Fortunately, training readers to precisely represent claim predicates does improve argument evaluation for many students (Larson et al., 2009).

In less-constrained situations, readers do not seem to use the criteria of relevance (or logical relatedness) as much as they use the criteria of acceptability (Shaw, 1996) or the criteria of whether or not the argument is consistent with one's belief (Wolfe et al., 2009). Shaw (1996), for instance, asked participants to produce objections to arguments selected from newspaper editorials. She found that readers objected to the acceptability of the claims and reasons embedded in the argument more often than they objected to the claim-reason relevance. Similarly, Wolfe et al. (2009) found that readers based their agreement judgments on the compatibility of their beliefs with the claims and reasons of arguments more than on the relevance connections. Participants read simple claim-reason arguments varied in terms of the side (pro/con) the claim asserted and the support-similarity (supporting-reason/opposing-reason) of the reason. Thus, each participant read arguments that were related to participants' beliefs as follows: congruent-claim/congruent-reason, incongruent-claim/incongruent-reason, congruent-claim/incongruent-reason, and incongruent-claim/congruent-reason. For each argument, readers, first rated their agreement with the argument, and, then, they rated the quality of the strength of the argument on 7-point Likert scales. Wolfe et al. (2009) found that readers made the agreement judgments based on the congruency of their belief with the claim while they made the quality evaluations on the congruency of their beliefs with the reason. In less-constrained situations, readers seem to base their agreement and quality judgments of the arguments on whether or not the individual elements of the arguments resonate with their beliefs, and not so much on the logical connection between them.

However, it is important to note that Shaw (1996) and Wolfe et al. (2009) did not directly test the readers' ability to base their evaluative judgments on relevance relations. For instance, one problem with the Shaw (1996) study findings is that she utilized complex, multi-sentence arguments as materials. We know from previous research (Britt & Larson, 2004) that undergraduates struggle to identify the claims and reasons, especially when these elements are embedded in multi-sentence complex arguments. Thus, the participants in the Shaw (1996) study may have had difficulty identifying the claims and reasons in the first place. As a result, readers could not object to the connections between claims and reasons. Instead, readers, when asked to produce objections, objected to the acceptability of the claims and reasons. Similarly, although the materials utilized in Wolfe et al. (2009) were not complex, and the evaluative task was more direct (agreement & quality) than the task in the Shaw (1996) study, the manipulation of the relevance of the claim-reason connections was not direct. Instead, Wolfe et al. (1996) directly manipulated whether the reason supported or opposed the claim, which was either in favor (pro) or not (con) of an asserted position. In other words, the arguments did not have simple unwarranted (reason simply is not relevant) arguments; the reasons provided rebuttals to the claims. Thus, the Shaw (1996) and Wolfe et al. (2009) studies did not directly test readers' ability to evaluate arguments based on the relevance criteria.

In a more direct test of students' skill in evaluating arguments, Larson et al. (2009) and Britt et al. (2008) manipulated argument quality by using unwarranted arguments (that is, arguments in which the "reasons" do not logically support the claim). These materials constitute a stronger test of readers' skill in using the criteria of relevance. They found that college students are only about 60-65% accurate when evaluating simple claim-reason arguments and improved

to only about 75% after a tutorial. In contrast, masters' students were about 90% accurate when evaluating these materials.

The Shaw (1996) and Wolfe et al. (2009) results, however, highlight the relationship between criteria of acceptability and argument evaluations. Both of these studies show that a reader, under less-constrained evaluative situations and while evaluating complex arguments or arguments that strikingly relate to the reader's beliefs, base their evaluative judgments on the acceptability criteria. By doing so, they rely on their beliefs and attitudes. Evaluations based on criteria of acceptability focus on the truth or accuracy of the claim and reasons. Since one typically regards one's own beliefs as accurate, evaluating the truth of a claim or reason is, in effect, a consistency check with one's own beliefs. Likewise, evaluating arguments based on personal agreement has been generally portrayed as the activation of associated attitudes on reading an argument (Voss et al., 1993). Such activation has been characterized as quick and, for strongly held attitudes, almost as fast as reading itself. It is, thus, not surprising that both Shaw and Wolfe et al. found acceptability judgments to be preferred over ones based on relevance. Indeed, Wolfe et al. (2009), who manipulated the sides (pro or con) of simple claim-reason arguments, found that readers based their agreement judgments largely on their agreement with the claim rather than on the strength of the reason.

People can employ different strategies when evaluating arguments. On tasks that require explicit discrimination between logically good and poor arguments, such as those used by Britt et al. (2008), some readers can read and represent claims precisely, thus enabling them to evaluate the relevance of reasons to the claim. On less stringent tasks, readers may focus only on the acceptability of the separate parts of the argument and on their congruence with existing beliefs and attitudes, rather than on precise logical relatedness.

Lexical Overlap

As the Britt et al (2008) results have shown, precision in memory for predicates of claims can affect whether or not readers base their evaluations on the logical relevance of the reasons. This implies that many lower level semantic and lexical aspects of the text could potentially affect readers' ability to evaluate the logical relatedness of claims and reasons because of their influence on memory structures. One such aspect is lexical overlap. Research with narratives has shown that lexical relatedness across text segments influences what prior information is reactivated during reading (O'Brien et al., 1995; O'Brien, Rizzella, Albrecht, & Halleran, 1998) as well as the time it takes to read and the amount of text recalled (Wolfe, Magliano, & Larson, 2005). In one particularly relevant example, Wolfe et al. (2005) manipulated the degree of lexical and causal overlap among pairs of sentences that described antecedents and consequences. For example, the consequence "He was carried unconscious to a hospital" was preceded by one of the following antecedents: "Malcolm collapsed while treating patients in his office" (high causal or high lexical overlap), "Malcolm was hit in the head with a stray baseball" (high causal or low lexical overlap), "Malcolm was a doctor in family practice" (low causal or high lexical overlap), or "Malcolm had a great seat for the baseball game" (low causal or low lexical overlap). Wolfe et al. (2005) found that lexical relatedness influences memory for narratives, particularly under lower levels of causality. Causal relatedness positively affects both memory (high causally related pairs remembered better than low causally related pairs) and processing (high causally related sentence pairs read faster than low related pairs) of narratives. Of more relevance for arguments, however, was that sentence pairs with high lexical overlap were recalled more frequently (though not read any faster) than low lexical overlap pairs, independent of causal relatedness.

Though related to improved recall, we cannot assume that the effect that lexical overlap has on argument evaluation is necessarily a positive one. Indeed, the Britt et al. (2008) findings showed a positive result only on memory for the precise claim predicate, something that is perhaps unlikely to be repeated in adjoining clauses in naturally occurring texts. This raises the question, however, of what effect overlap of other parts of the claim might have on evaluation. One suggestion can be drawn from work on the detection of semantic inconsistencies. Barton and Sanford (1993), for instance, gave participants questions containing inconsistencies with either high lexical overlap (for example, "When an aircraft crashes, where should the survivors be buried?") or low lexical overlap (for example, "When a bicycle accident occurs, where should the survivors be buried?"). They found that high lexical overlap (air crashes and death) led to fewer detections of the inconsistency (one does not bury survivors) than did low overlap (bike crashes and death). Thus, lexical overlap could make it more difficult to evaluate the consistency of two clauses.

Research also suggests that the effect of lexical semantics may be even more pronounced for less-skilled readers (Hannon & Daneman, 2004; Perfetti, Goldman, & Hogaboam, 1979; Todaro, Millis, & Dandotkar, 2010). For example, Hannon and Daneman (2004), using the Barton and Sanford (1993) materials, found that less-skilled readers were more affected by lexical semantics than were skilled readers. Likewise, Todaro et al. (2010), using the Wolfe et al. (2005) materials, also found that less-skilled readers were more affected by lexical overlap than were skilled readers. They had participants judge how coherent the stories were. They found a three-way interaction between reading skill, lexical overlap, and causal overlap. When there was a low level of causal relatedness, both high and low skilled readers judged high lexical overlap pairs as more coherent than low overlap pairs. However, when there was high causal relatedness,

judgments from less-skilled readers were significantly more affected by the degree of lexical overlap than were judgments from skilled readers. These results suggest that high skilled readers give higher priority to deep semantic relationships than to low semantic relationships. Conversely, low skilled readers are affected by low-level relationships regardless of the strength of high-level relationships, suggesting a greater inclination for shallow processing.

Lexical overlap may affect argument evaluation through its effect on a reader's memory of argument elements. This effect, however, may not be positive if the overlap reduces focus on the claim predicate or otherwise induces a more gist-like representation. Furthermore, less-skilled readers appear to be more affected by lexical influences than are skilled readers. Thus, if the effect of lexical relatedness on argument evaluation were negative, we might expect it to be more pronounced with less-skilled readers.

Overview of Experiments

We conducted two experiments that independently manipulated the logical and semantic relatedness between claims and reasons of arguments such as 1a-1d. Participants read reason-claim ordered arguments one-clause-at-a-time and made evaluative judgments. The arguments were presented one-clause-at-a-time, to emphasize the distinct argument element for the evaluation task and so that the reason could be used as a cue for the claim recall task. The arguments were presented in the reason-claim order, so that participants read an identical target region (i.e., the claim) in all conditions (See Table 1 for an example item). In Experiment 1, evaluative instructions emphasized evaluation of the relevance of the claim-reason relationship by asking participants to decide whether the arguments were structurally acceptable or flawed, without considering their opinion. In Experiment 2, evaluative instructions stressed the claim-attitude relationship by asking participants to rate their agreement with the arguments. After

evaluating all the arguments, participants were asked to recall the arguments. In each experiment, we examined the extent to which evaluations were based on logical and semantic relatedness and whether skilled and less-skilled reasoners differentially rely on these two potential sources of coherence.

Experiment 1

The essential coherence relationship for arguments is logical relatedness among an argument's claims and reasons (Angell, 1964; Blair & Johnson, 1987; Freeman, 1991; Salmon, 1984; Shaw, 1996; Voss & Means, 1991). Thus, argument evaluation should be guided primarily by verifying logical relatedness. Prior research has also demonstrated the importance of logical relatedness on judgments of the quality (Wolfe et al., 2009) and memory of an argument (Britt et al., 2008; Larson et al., 2009). Therefore, we anticipated that there would be a main effect of logical relatedness on quality evaluations. Moreover, based on Larson et al. (2009), less-skilled reasoners should not be as skilled in knowing how to apply the criteria of relevance. Therefore, there should be an interaction of logical relatedness by skill such that the effect of logical relatedness will be greater for skilled reasoners than it is for less-skilled reasoners. Finally, based on prior research (Larson et al., 2009; Britt et al., 2008), there should be a main effect of reasoning skill on recall such that skilled reasoners should recall predicates more accurately than less-skilled reasoners.

The role of lexical overlap is more complicated. Predictions can be made by extrapolating from the findings on how skilled and less-skilled readers process lexical overlap in the context of other forms of discourse. Less-skilled readers are more prone to shallow processing (Hannon & Daneman, 2004) and more affected by lower level semantic relationships (Perfetti et al., 1979). Therefore, as with the findings of Todaro et al., (2010), it is reasonable to predict that there

would be a reasoning skill by lexical overlap interaction such that the effect of lexical relatedness on evaluation will be greater for less-skilled reasoners than it is for skilled reasoners. Skilled reasoners' evaluations should be primarily affected by logical relatedness.

Method

Participants. Participants were forty-three undergraduates from Northern Illinois University who were prescreened for the reasoning skill based on their performance on the verbal reasoning test described below.

Design and materials. A 2 (Logical Relatedness: High Logical vs. Low Logical) X 2 (Lexical Overlap: High Semantic vs. Low Semantic) X 2 (Reasoning Skill: Skilled vs. Less-skilled) mixed factorial design, with reasoning skill as the between-participant factor, and logical relatedness and lexical overlap as the within-participant factors was adopted. Twenty-four claims of policy were created with topics that were not expected to evoke strong emotions or attitudes on the part of the students (for example, abortion) but would be somewhat engaging for the population. The claims did not contain negations or low frequency predicates (for example, deter, abolish). Reasons were created by starting with a high logical reason (relevant support for the claim) and a low logical reason (a true descriptive statement that would not support the believability of the specific claim predicate. These versions were considered the high lexical overlap condition (see Table 1 for an example argument). Following Kurby, Britt, & Magliano (2005), the low lexical overlap condition was created by replacing specific content words in the high lexical items with less specified, but semantically consistent words (for example, 'kids' replaced with 'humans', 'infectious' replaced with 'serious'). In this way, the logical claim-reason relationship was the same for the high and low semantic versions, but there would be fewer shared lexical features. We equated the number of syllables in the reasons across

conditions (High Logical or High Semantic: *M* = 12.38, SD = 1.69; High Logical or Low Semantic: *M* = 11.88, *SD* = 1.51; Low Logical or High Semantic: *M* = 12.08, *SD* = 1.59; Low Logical or Low Semantic: *M* = 11.54, *SD* = 1.79). Four lists of arguments were created using a Latin-square method to assign the four reasons in each group to the twenty-four claims. Each participant received six items in each condition and only one version of each reason for the given claim.

The manipulation of lexical overlap was confirmed by a norming study. Each participant was presented with content words from the claims and reasons of arguments in two different columns. Thirty-eight participants rated the semantic relatedness between the groups of words related to a claim and the reason on a 6-point scale (1 = *Not at all related*; 6 = *Very related*). A 2 Lexical Relatedness (High vs. Low) X 2 Logical Relatedness (High vs. Low) ANOVA on perceived semantic-overlap scores confirmed the manipulation, yielding a significant effect of semantic relatedness ($F1(1, 37) = 105.47$, $MSE = .58$, $p < .001$, $\eta 2 = .45$; $F2(1, 23) = 118.89$, $MSE = .33$, $p < .001$, $\eta 2 = .49$). Indeed, participants rated the high lexical items ($M = 4.54$; $SE = .09$) as more semantically related than the low lexical items ($M = 3.27$; $SE = .11$). A series of simple effect tests found that the participants rated high lexical items as more semantically related than low lexical items for both high logical, $t1(37) = 7.21$, $p < .001$; $t2(23) = 8.27$, $p < .001$, and low logical $t1(37) = 9.22$, $p < .001$; $t2(23) = 6.95$, $p < .001$, arguments. No other findings were significant, suggesting that the semantic manipulation was independent of the logical connection between claims and reasons.

Verbal reasoning test. Participants were preselected from a large pool of 531 students based on their reasoning skill, which was determined by their performance on the verbal reasoning section of the Law School Admissions Test (LSAT). The verbal reasoning test was

administered as a timed task (20 minutes), in groups of around twenty participants in a classroom, as a part of the mass testing that was done in the weeks prior to the experiment. The verbal reasoning test consisted of fourteen short argument texts each followed by one or two multiple-choice questions that tested readers' abilities to identify argument elements, such as the claim or a key assumption, to identify one of a set of common flaws or to identify a way to strengthen or weaken the author's argument. Overall, participants' scores on the test were low, as expected (M= 7.72 of 18 possible). Participants were classified as less-skilled when their scores fell in a range of zero to five and as skilled reasoners when their scores were ten or above. The average scores for twenty-two skilled reasoners (M = 11.73, SD = 1.4) and nineteen less-skilled reasoners (M = 3.7, SD = 1.5) were significantly different from each other, $t(39) = 17.45$, $p<.001$.

Procedure. Participants evaluated four practice arguments to become familiar with the task and then their randomly assigned list of reason-claim arguments. For each argument, the reason was presented first, and the participants pressed the spacebar to reveal the claim and remove the reason. With the claim on the screen, participants were asked to decide whether the argument was an *acceptable* (Ok) or *flawed* argument (Flawed), given only the structure of the argument, not how convincing it was and irrespective of whether they agreed with it or not. Participants pressed the "F-labeled" key (the 'F' key on the keyboard) for the structurally flawed judgment, and they pressed the "O-labeled key" (the 'J' key on the keyboard) for the structurally acceptable judgment. Participants were instructed to make quality-evaluative judgments as fast as they could and to keep their fingers on the "F" and "O" labeled keys throughout the experiment.

After evaluating the twenty-four arguments, participants completed a surprise, cued-recall task in which the themes of the argument claims (for example, vaccinations in Argument

1a) were presented, one theme at a time, on the computer screen. Participants were instructed to recall the related arguments in as close to the exact words as possible.

Scoring. One skilled and one less-skilled reasoner's data were dropped from the analyses because their recall responses were blank. A 2 X 2 X 2 (Lexical Overlap X Logical Relatedness X Reasoning Skill) analysis of variance was conducted on quality judgments and predicate-recall score. Quality judgments were defined as the proportion of arguments in a condition that the participants evaluated as acceptable. The cued-recall task was scored for predicate-recall accuracy. Predicates in the claim were scored as correct if they were exact matches (for example, "should be required" for Arguments 1 a-d) or close synonyms (for example, "should be mandated"). The coding of predicate-recall scores had a high inter-rater reliability (k =.90).

Both quality judgments and predicate-recall accuracy were arcsine transformed because they were proportional scores with binomial forms (Accurate or Inaccurate) of distribution (Kirk, 1982). Participant and item analyses, referred to as F_1 and F_2 respectively, were conducted for each measure. All follow up tests were analyzed using Bonferroni corrections.

Results

Quality judgments. Average quality judgment scores as a function of lexical overlap and logical relatedness for skilled and less-skilled reasoners are displayed in Table 2. There were statistically reliable main effects of logical relatedness ($F_1(1, 39) = 104.81$, $MSE = .09$, $p < .001$, $\eta^2 = .69$; $F_2(1, 46) = 101.65$, $MSE = .12$, $p < .001$, $\eta^2 = .69$) and lexical overlap ($F_1(1, 39) = 8.14$, $MSE = .08$, $p < .01$, $\eta^2 = .17$; $F_2(1, 46) = 10.21$, $MSE = .07$, $p = .003$, $\eta^2 = .18$). Both main effects indicated that judgments were higher for items high on the dimensions than for items low on the dimensions. However, these main effects were qualified by logical relatedness X lexical overlap interaction that was statistically significant for the participant, $F_1(1, 39) = 6.58$, $MSE = .04$, p

=.014, η^2 =.14, but not for the item analysis, $F_2(1, 46) = 2.33$, *MSE* =.08, p =.13, η^2 =.05. We conducted two follow up t-tests to further explore the nature of the interactions shown in Figure 1. With respect to high logical arguments, the judgments for high lexical items were higher than the judgments for low lexical items, $t(40) = 3.25$, p =.002; d = 1.03. In contrast, for the low logical arguments, there was no such difference between high and low lexical overlap, $t(40)$ =.85, p =.40; d =.27.

With respect to the impact of reasoning skill, the main effect of skill was not significant, (F=2.3, p=.138). However, the interaction between reasoning skill and logical relatedness was significant,$F_1(1, 39) = 8.95$, *MSE* =.09, p =.005, η^2 =.06; $F_2(1, 46) = 6.71$, *MSE* =.12, p =.013, η^2 =.04. A follow up analysis to further explore the interaction, shown in Figure 2, suggested that the effect of logical relatedness was significantly higher for skilled (*M* difference =.64) than for less-skilled reasoners (*M* difference =.35). No other results were significant.

Predicate recall. The means for predicate-recall scores are presented in Table 3. There was a significant main effect of reasoning skill, $F_1(1, 39) = 15.15$, *MSE* =.18, p <.001, η^2 =.33; $F_2(1, 46) = 13.12$, *MSE* =.30, p =.001, η^2 =.22. As expected, skilled reasoners (*M* =.71, *SE* =.05) recalled the predicates better than the less-skilled reasoners (*M* =.45, *SE* =.05). Additionally, the interaction between reasoning skill and logical relatedness was significant for item analysis, $F_2(1, 46) = 4.32$, *MSE* =.07, p =.043, η^2 =.08, but not for participant analysis, $F_1(1, 39) = 1.53$, *MSE* =.13, p =.29, η^2 =.33.

Discussion

Consistent with the predicted results regarding logical relatedness, both skilled and less-skilled reasoners made higher number of quality judgments for high logically related arguments than for low logically related arguments, and as expected, the difference between the two was

larger for high skilled reasoners than for low skilled reasoners (See also, Larson et al., 2009). Lexical overlap also influenced quality judgments but not as predicted. Specifically, high lexical overlap led to higher quality judgments than low lexical overlap did, but only when there was high logical overlap. This stands in contrast to Todaro et al. (2010) who found that lexical overlap affects judgments of coherence of simple narratives only for less-skilled readers. This suggests that the process of evaluating the quality of an argument is different from the process of evaluating the coherence of a narrative. In the former, perceived lexical overlap may facilitate ones' ability to identify the presence of a logical relationship between a claim and a reason. There are two potential mechanisms for why lexical overlap influences quality judgments. First, the degree of lexical overlap should affect the activation of the reason upon reading the claim (Kintsch, 1988; 1998; Myers & O'Brien, 2005), and, second, it should affect memory-based integrative processes (Kintsch, 1988; 1998). However, there must be a sufficient level of perceived logical relatedness for lexical overlap to affect judgments.

The results of Experiment 1 suggest that lexical overlap can facilitate, rather than hinder, the evaluation of the quality of arguments. Difficulty in evaluating the quality of arguments may likely be due to imprecision in recalling the predicates of argument claims (as shown by the skill effect on claim memory), as well as to a lack of sufficient sensitivity to the logical connections (as shown by the skill interaction on logical relatedness). However, it is important to examine the relative effects of these two factors on a less-directed test. The quality evaluation task required reasoners to attend specifically to the structure of the argument. In Experiment 2, we examine the spontaneous reliance on the criterion of logical relatedness when the task allows for additional criteria, including evaluation based on one's level of agreement with an argument.

Experiment 2

Experiment 2 examined the spontaneous reliance on the criteria of logical and lexical relatedness when making judgments of agreement. Prior research has shown that readers based agreement judgments relatively more on an evaluation of the claim's acceptability than when making quality judgments, which were more affected by the type of reason (Wolfe et al., 2009). This experiment, however, did not include reasons that were not relevant to making an evaluation of the logical connection between the claim and the reason. Thus, it is unclear how sensitive readers are to logical relatedness when making agreement judgments. While there may be other factors that influence whether or not a reader agrees with an argument, the argument should be less persuasive if the reason does not provide support, as in the low logical cases. Thus, it is predicted that there will be a main effect of logical relatedness even for agreement judgments.

Method

The material and procedure used in Experiment 2 are similar to those used in Experiment 1, except that participants in this experiment were asked to rate their agreement with arguments based on a 5-point scale. Similar to the first experiment, the critical independent variables for the proposed experiment were logical (high or low) and semantic (high or low) relationships between claims and reasons of arguments and reasoning skill (skilled/less-skilled). Similarly, the dependent measures for the current experiment are agreement judgment scores and predicate-recall score.

Participants. Participants were sixty-two undergraduates from Northern Illinois University who, similar to Experiment 1, were prescreened for reasoning skill. The average score on the LSAT of skilled (>=10; M = 11.21) and less-skilled (<=5; M = 4.24) reasoners differed

significantly, $t(60) = 17.99$, $p < .001$.

Procedure and design. Participants were presented with the same arguments as in Experiment 1. However, in this experiment, participants rated their agreement with the argument on a 5-point scale (1 = *Strongly Agree*; 2 = *Agree*; 3 = *Neutral*; 4 = *Disagree*; 5 = *Strongly Disagree*). Similar to Experiment 1, a 2 (Lexical Relatedness: High lexical vs. Low lexical) X 2 (Logical Relatedness: High Logical vs. Low Logical) X 2 (Reasoning skill: Skilled vs. Less-skilled) mixed factorial design with reasoning skill as a between-participant factor and semantic relatedness and logical relatedness as within-participant factors, was adopted.

Scoring. No participant's data were dropped from any of the analyses. In order to ease interpretation of the means, the participant responses for the agreement judgments were flipped (1 = *Strongly Disagree*; 2 = *Disagree*; 3 = *Neutral*; 4 = *Agree*; 5 = *Strongly Agree*) so that higher numbers referred to agreement and lower numbers referred to disagreement**.** Mean agreement judgment scores were calculated for each participant and condition (F_1) and for each item and condition (F_2).

Similar to Experiment 1, recall responses from Experiment 2 were coded for claim-predicate accuracy. The coding of recall responses for predicate-recall accuracy had a high inter-rater reliability ($k = .92$). Similar to Experiment 1, the proportion of predicates accurately recalled was calculated for each participant and condition, and the scores were transformed using arcsine (Kirk, 1982).

Results

Agreement judgments. Agreement judgment score for each condition and skill are presented in Table 2. A three-way ANOVA revealed a significant main effect of logical relatedness, $F_1(1, 55) = 43.64$, $MSE = .27$, $p < .001$, $\eta^2 = .44$; $F_2(1, 46) = 35.71$, $MSE = .28$, p

<.001, η^2 =.44. High logical arguments (*M* = 3.36, *SE* =.05) had a higher agreement rating than low logical arguments (*M* = 2.90, *SE* =.06). No other findings were significant.

Predicate recall. The predicate-recall score for each condition is presented in Table 3. The analysis revealed a significant main effect of reasoning skill, $F_1(1, 55) = 5.27$, *MSE* =.23, *p* =.026, η^2 =.09; $F_2(1, 46) = 7.41$, *MSE* =.31, *p* =.009, η^2 =.14, in that skilled reasoners' (*M* = 0.67, *SE* =.05) recall was better than that of less-skilled reasoners' (*M* = 0.53, *SE* =.04). There was also a significant interaction between reasoning skill and logical relatedness, $F_1(1, 55) = 4.77$, *MSE* =.12, *p*=.033, η^2 =.08; $F_2(1, 46) = 8.59$, *MSE* =.06, *p* =.005, η^2 =.16. A follow up analysis to further explore the interaction revealed that skilled reasoners' difference in predicate-recall scores between high (*M* =.72, *SE* =.06) and low (*M* =.63, *SE* =.06) logical arguments did not differ, $t(25) = -1.12$, *p*=.27; *d* =.45, whereas less-skilled reasoners had better memory for high logical (*M* =.59, *SE* =.05) than low logical predicates (*M* =.47, *SE* =.05), $t(30) = 2.08$, *p*=.046; *d* =.76.

Discussion

Experiment 2 found that readers spontaneously used logical relatedness as a factor in evaluating agreement with arguments and, unlike in Experiment 1, there was no difference in the magnitude of this effect based on skill. Both skilled and less-skilled reasoners were equally sensitive to the claim-reason connection when evaluating the strength of their agreement with arguments. Although skilled reasoners are more accurate at distinguishing relevant from irrelevant reasons, less-skilled reasoners still use this information when judging arguments, even when this is not the primary task. Thus, at least for arguments that are not too controversial, readers do consider whether or not the reason is relevant when judging the strength of their agreement with the argument. The only skill effect was on predicate-recall. As in Experiment 1,

less-skilled reasoners had worse recall for the claims overall than did skilled reasoners, and they had especially poor recall for low logical arguments as compared to high-logical arguments. Skilled reasoners, on the other hand, recalled predicates equally well across high and low-logical conditions. Also, overall, readers did not differ in their recall of high and low-logical predicates. This lack of an effect of logical relatedness on recall runs counter to what has typically been shown in the context of narrative comprehension, whereas recall tends to be higher for texts that have a high level of deep semantic relationships (e.g., causal) than for those that do not (for example, Myers, Shinjo & Duffy, 1987; Trabasso & van den Broek, 1985; Wolfe et al., 2005).

The findings also show that readers rely on logical relatedness more than on lexical overlap, even under agreement tasks. Unlike Experiment 1, semantic overlap did not affect agreement judgments. Thus, readers are not fooled into accepting arguments simply because they are related lexically. The logical relationship is what matters most.

General Discussion

These experiments allowed us to independently test the impact of logical and lexical relatedness on two evaluation tasks that differ in terms of task directness. Several important conclusions can be made from these experiments. First, logical relatedness guides the evaluation of both types of judgments. Clearly, logical relatedness should govern quality judgments in which the primary task is an analytical evaluation of logical relatedness, which has been previously demonstrated (e.g., Britt et al., 2008). For this task, we replicated the effect that reasoners can target their evaluation to this criterion and that those who have lower general reasoning skill are also less able to target their application of this relevance criteria (Larson et al., 2009). However, it was surprising that agreement judgments were primarily affected by logical relatedness even though this task does not explicitly prompt the use of this criterion.

Second, Experiment 2 replicated the finding that those skilled at accurately recalling the predicates of claims were also more accurate at making quality judgments (Britt et al., 2008) and extended this effect of memory precision on a quality judgment task (Experiment 1) and after a delay. The relationship between recall and reasoning is consistent with Kyllonen and Christal's (1990) experiments that showed high correlations between working memory and a variety of deductive and inductive reasoning tasks. Others have also found high correlations between working memory and the Raven's progressive matrices reasoning test (Ackerman, Beier, & Boyle, 2005; Engle, Tuholski, Laughlin, & Conway, 1999). These results further support helping less skilled reasoners attention to and retrieve claim predicates when evaluating arguments (Larson et al., 2009).

Third, shallow semantics helped under some circumstances, but, contrary to predictions, they did not dupe less-skilled reasoners into thinking that the arguments are stronger than they are. This prediction was based on prior research that has shown that various aspects of comprehension for less skilled readers are more affected by shallow semantics than skilled readers (Hannon & Daneman, 2004; Perfetti et al., 1979; Todaro et al., 2010). However, the results of Experiment 1 suggest that lexical overlap can actually facilitate, rather than hinder, the evaluation of the quality of arguments for both skilled and less-skilled reasoners. Specifically, high lexical overlap led to higher quality judgments than low lexical overlap did, but only when there was high logical overlap. Why might this be the case? Evaluating argument quality requires an analysis of the warrant or general principle relating the reason to the claim. When there is low logical overlap, the two elements of the argument will have a general principle that does not make sense (for example, that we should require children to take weakened things). In contrast to high logical arguments, where the general principle makes sense even if the reader does not

agree with it (for example, that we should require children to take things that protect them from harm), this all depends on having an accurate representation of the argument. The key to creating a coherent representation is to have a precise representation for the predicate of the claim (Britt et al., 2008). In particular, the reader has to represent the reason, represent the claim, and then infer the general principle between these elements (the warrant). One possible explanation for how lexical overlap can enhance argument evaluation is it may serve as a retrieval cue upon reading the claim to reactivate the previously read reason. When there is low logical relatedness, the resultant warrant, or connection, is easily evaluated as nonsensical. In the high logical arguments, however, these warrants are not immediately rejected, and the decision that must be made is whether or not it exceeds some minimally sufficient level. At this point, high levels of lexical overlap should facilitate the process of constraint satisfaction that governs integration (Kintsch, 1988; 1998). Increasing successful integration can, therefore, lead to more accurate judgments for warrants that are minimally supportive of the claim.

On the other hand, shallow semantics do not appear to affect judgments of agreement. Of course, many other factors will likely impact agreement judgments, such as beliefs and attitudes toward the claim (Voss et al., 1993; Wolfe et al., 2009), degree of prior knowledge associated with the topic of the argument (Larson, Britt, & Kurby, 2009), and the nature of the reason or principle connecting them (Shaw, 1996). However, it is interesting to note that readers based their agreement judgments on logical relatedness, even when the task did not require them do so. At the same time, the logical connections that readers based their agreement judgments on were not facilitated by semantic overlap as they were when the readers were evaluating quality judgments (in Experiment 1). Given the less-constrained situation of making agreement judgments, it is possible that readers base these agreement judgments on a dynamic interaction

among various factors, including, beliefs and logical relatedness. Certainly, a more definitive assessment of how beliefs and prior knowledge may interact with lexical and logical relationships in arguments would require further research.

Fourth, these experiments suggest that argument quality and agreement judgments are different from coherence judgments. That is, Todaro et al. (2010) found that readers used semantics as an additional cue to signal when relatedness was low. However, their task was to judge whether two clauses "fit together." It seems that the standards for quality and agreement judgments are more clearly logic-based than general-relatedness based. One general conclusion that can be drawn from the differences between the present study and that of Todaro et al. (2010) is the processes involved in understanding and evaluating arguments may be different from those of narratives. The results of this study, in contrast to those of Todaro et al. (2010), are consistent with research that shows that genre expectations affect discourse processing (for example, Zwaan, 1996).

One may argue that the difference in the current and narrative study findings may be because of the difference in the way the deeper coherence relationship was manipulated. We presented arguments with the connector, 'because,' which may have signaled the logical connection to the reader. This signaling might have primed the readers to anticipate a logical connection even before reading the claim. This, however, was not the case with narrative studies. For instance, Todaro et al. (2010) presented sentence pairs that varied in content and had no connectors. Yet, the current study findings still count and are equally informative in that there was a clear effect of logical relatedness in both of the evaluative tasks, despite the signaling. If anything, this difference underscores how much of an effect the genre may have on processing discourse.

One practical implication of these findings is that the fear that people are excessively sensitive to spurious arguments that are based on relationships that rest on lower level semantics may not be warranted. Readers presented with arguments that vary in logical adequacy will use the criteria of relevance, and they do so even under more spontaneous conditions. The results suggest that shallow processing of the claim, but not shallow evaluation of the connection, seems to be more important for understanding why less-skilled reasoners struggle. In fact, it may help less-skilled reasoners to use lexical overlap during the evaluation process.

Nevertheless, it is important to consider the limitations of these studies when applying the current study findings to real life. It is possible that the difference in the effect of semantic overlap between the first and second experiments is due to the difference in the response-scale that Experiments 1 and 2 adopted. Whereas, a forced choice (flawed or acceptable) response was adopted for quality judgments, a five-point likert-scale response was used for the agreement judgments. Although a likert-scale was more appropriate for the purposes of capturing agreement responses because of the less-constrained nature of the task, it is still possible that the participants might have had stronger opinions about some of the topics despite the fact that we took measures to avoid topics that people might have strong opinions about (for example, abortion). Furthermore, although we took care to ensure that the reasons were equally plausible across the manipulated conditions, it is possible that the items might be more probable in one condition than in the others. Future studies with material controlled for the opinion strength of arguments and plausibility of reasons across conditions would give us more information about the issue.

Conclusion

The current study findings have implications on argumentation research and on educational practices. First, not only have we found that the factors related to relevance connections in an argumentative text influence how one evaluates arguments, but we have also found that these factors, at least the coherence relations kind, affect the evaluations of the argumentative and non-argumentative texts differently. Second, the current study findings corroborate our earlier finding that precisely retaining the predicates is essential for evaluating arguments, suggesting how important it is for educators to incorporate this technique in instructions. Finally, we now know that excessively relying on word overlap, which less-skilled readers do, is not why some struggle to evaluate arguments. Instead, precision in representing claim predicates is what makes a reader skilled at evaluating arguments. Thus, one's ability to evaluate arguments, which is crucial to academic literacy, is more complex than being able to read well, and there is a need for more research on arguments as texts to better understand how we evaluate arguments and how we can inform educators to instill these skills in students.

References

Achieve, Inc. (2013). Next generation science standards. Washington, DC: Author.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*, 30-60.

Angell, R. B. (1964). *Reasoning and logic*. New York: Appleton-Century-Crofts.

Britt, M. A., Kurby, C. A., Dandotkar, S., & Wolfe, C. R. (2008). I agreed with what? Memory for simple argument claims. *Discourse Processes, 45*(1), 52-84. doi: 10.1080/01638530701739207

Britt, M.A, & Larson, A. A. (2003). Constructing representations of arguments. *Journal of Memory and Language, 48*, 794-810. doi: 10.1016/S0749-596X(03)00002-0

Copeland, D.E., & Radvansky, G.A. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology, 57A*(8), 1437-1457. doi: 10.1080/02724980343000846

Council of Chief State School Officers. (2010). The common core standards for English language arts and literacy in history/social studies and science and technical subjects. Washington, DC: National Governors Association for Best Practices. Retrieved from http://www.corestandards.org

Durik, A. M., Storey, J. K., Kopp, K., & Britt, M.A. (May 2008). *The Combined Effects of Motivational and Cognitive Factors in Predicting Success on a Learning Task.* Poster presented at the Society for the Study of Motivation. Chicago, IL.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: a latent variable approach. *Journal of Experimental Psychology: General, 128*, 309–331.

Fisher, A. (1988). *The logic of real arguments.* Cambridge, England: Cambridge University Press.

Freeman, J. B. (1991). *Dialectics and the macrostructure of arguments: A theory of argument structure.* Berlin: Foris Publications.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Graesser, A., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371-395. doi:10.1037/0033-295X.101.3.371

Hannon, B., & Daneman, B. (2004). Shallow semantic processing of text: An individual-differences account. *Discourse Processes, 37,* 187-204. doi: 10.1207/s15326950dp3703_1

Johnson, R. H., & Blair, J. A. (1977). *Logical self-defense*. Toronto: McGraw-Hill Ryerson.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182. doi:10.1037/0033-295X.95.2.163

Kintsch, W., van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394. doi: 10.1037/0033-295X.85.5.363

Kirk, R. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed). Belmont, CA: Brooks/Cole.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77*, 319-337. doi: 10.1002/sce.3730770306

Kurby, C. A., Britt, M. A., & Magliano, J. P. (2005). The role of top-down and bottom-up processes in between-text integration. *Reading Psychology, 26*(4-5), 335-362. doi:10.1080/02702710500285870

Kyllonen, P., & Christal, R. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence, 14*, 389-433.

Larson, A. A., Britt, M. A., & Kurby, C. A. (2009). Improving students' evaluation of informal arguments. *The Journal of Experimental Education, 77*(4), 339-365. doi:10.3200/JEXE.77.4.339-366

Long, D. L., & Lea, R. B. (2005). Have been searching for meaning in all the wrong places? Defines the "search after meaning" principles in comprehension. *Discourse Processes, 39*(2&3), 279-298. doi:10.1080/0163853X.2005.9651684

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098-2109. doi: 10.1037/0022-3514.37.11.2098

MacDonald, M. C., Just, M., A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognition, 24*, 56-98. doi: 10.1016/0010-0285(92)90003-K

McNamara, D.S., Louwerse, M.M., McCarthy, P.M., & Graesser, A.C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*, 292-330. doi: 10.1080/01638530902959943

National Assessment of Educational Progress (2013). *NAEP* 2013 *U.S. Writing Report Card: Findings from the National Assessment of Educational Progress.* Princeton, NJ: Educational Testing Service.

National Center for History in Schools. (1996). *National Standards for History.* (Basic Ed.). Los Angeles: Author.

National Science Board, 2006: *America's pressing challenge-building a stronger foundation.* Retrieved July 20, 2006, from http://www.nsf.gov/statistics/nsb0602/nsb0602.pdf

National Science Education Standards. *An Overview*. (1996). Retrieved June 5, 2012, from http://darwin.nap.edu/html/nses/overview.html

No Child Left Behind Act of 2001. Pub. L. No. 107-110 (2002). doi: 10.1525/jer.2007.2.1.101a

Perfetti, C., Goldman, S. R., & Hogaboam, T. (1979). Reading skill and the identification of words in discourse context. Memory and Cognition, 7, 273–282. doi: 10.3758/BF03197600

Perkins, D. N., Faraday, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 83-106). Hillsdale, NJ: Erlbaum.

Salmon, M. H. (1984). *Logic and critical thinking*. Orlando, FL: Harcourt Brace Jovanovich.

Shaw, V. F. (1996). The cognitive processes in informal reasoning. *Thinking & Reasoning, 2*(1), 51-80. doi:10.1080/135467896394564

Todaro, S. A., Millis, K. K., & Dandotkar, S. (2010). The impact of semantic and causal relatedness and reading skill on standards of coherence. *Discourse Processes, 47,* 421-446. doi:10.1080/01638530903253825

Toulmin, S. E. (1958). *The uses of argument*. Cambridge, MA: Cambridge University Press.

Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes, 21*(3), 255-287. doi:10.1080/01638539609544959

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24*, 612-630. doi: 10.1016/0749-596X(85)90049-X

Verschuerem, N., Schaeken, W., & D'Ydewalle, G. (2005). Everyday conditional reasoning: A working memory-dependent tradeoff between counterexample and likelihood use. *Memory and Cognition, 33*(1), 107-119. doi: 10.3758/BF03195301

Voss, J. F. (2006). Toulmin's model and the solving of ill-structured problems. *Argumentation, 19*, 321-329. doi: 10.1007/978-1-4020-4938-5

Voss, J. F., & Means, M. L. (1991). Learning to reason via instruction in argumentation. *Learning & Instruction, 1*(4), 1991, 337-350. doi:10.1016/0959-4752(91)90013-X

Voss, J. F., Fincher-Kiefer, R., Wiley, J., & Silfies, L. N. (1993). On the processing of arguments. *Argumentation, 7*(2), 165-181. doi:10.1007/BF00710663

Voss, J. F., Wiley, J., & Carretero, M. (1995). Acquiring intellectual skills. *Annual Review of Psychology, 46*(1), 155-181. doi: 10.1146/annurev.psych.46.1.155

Wiley, J. (2005). A fair and balanced look at the news: What affects memory for controversial arguments? *Journal of Memory and Language, 53*, 95-109. doi: 10.1016/j.jml.2005.02.001

Wolfe, C. R. (2011). Argumentation across the curriculum. *Written Communication, 28*, 193-219. doi:10.1177/0741088311399236

Wolfe, C. R., & Britt, A. M. (2008). The locus of my side bias in written argumentation. *Thinking and Reasoning, 14*(1), 1-27. doi: 10.1080/13546780701527674

Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication, 26,* 183-209. doi: 10.1177/0741088309333019

Wolfe, M. B. W., Magliano, J. P., & Larsen, B. (2005). Causal and semantic relatedness in discourse understanding and representation. *Discourse Processes, 39*(2&3), 165-187. doi:10.1207/s15326950dp3902&3_4

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185. doi:10.1037/0033-2909.123.2.16