Jahrgang 30 / Heft 2–3 / 2016

Zeitschrift für

Pädagogische Psychologie

German Journal of Educational Psychology

Themenheft/Special Issue

Learning through Communication

hogrefe

Inhalt

Zeitschrift für

Pädagogische Psychologie

hogrefe

Original Article

Identifying General Cognitive Abilities Involved in Argument Comprehension and Evaluation

M. Anne Britt[1], Kristopher J. Kopp[1], Amanda M. Durik[1], Dylan Blaum[1], and Peter Hastings[2]

[1] Northern Illinois University, DeKalb, IL, USA
[2] DePaul University, Chicago, IL, USA

Abstract: Individuals who lack the necessary skills for accurate argument evaluation will be at a significant disadvantage throughout their lives. Research has shown that simple tutorials are not effective for many students (about 30 %). The current research examines how general abilities (e.g., vocabulary knowledge, reading comprehension, and analytical reasoning) predict learning to evaluate arguments. Study 1 showed that, although all three cognitive abilities predicted some aspects of argument comprehension and evaluation skills, vocabulary knowledge positively predicted both baseline argumentation skills and improvement from exposure to a tutorial. Study 2 found that this relationship between vocabulary knowledge and argument evaluation skill is partially mediated by general vocabulary ability and cannot be accounted for by knowledge of the meaning of the specific predicates used in the argument task. The results suggest that targeting skilled use of vocabulary knowledge and overall lexical quality may help students who do not learn from a simple tutorial.

Keywords: argument comprehension, argument evaluation, learning, vocabulary, reading and reasoning skill

Das Verständnis und die Evaluation von Argumenten: Zur Rolle allgemeiner kognitiver Fähigkeiten

Zusammenfassung: Personen, denen die notwendigen Fähigkeiten für die korrekte Bewertung von Argumenten fehlen, ziehen ein Leben lang daraus Nachteile. Forschungsbefunde zeigen, dass einfache Tutorien für viele Studierende (ca. 30 %) keinen Nutzen bringen. Wir berichten Daten zur Frage, wie sich durch allgemeine Fähigkeiten (z. B. Vokabelwissen, Leseverständnis, analytisches Schlussfolgern) das Erlernen der Argumentevaluation vorhersagen lässt. In Studie 1 konnte gezeigt werden, dass – obwohl alle drei kognitiven Fähigkeiten einige Aspekte von Argumentverständnis und Evaluationsfähigkeit vorhersagen – das Vokabelwissen sowohl die Argumentationsfähigkeit und nicht durch das Wissen um die Bedeutung einer spezifischen Aussage, die in der Aufgabe gestellt wurde, erklärt. Die Ergebnisse zeigen auf, dass der gezielte Einsatz von Vokabelwissen und allgemeiner lexikalischer Qualität diesen Studierenden, die nicht von einem einfachen Tutorium profitieren, helfen könnte.

Schlüsselwörter: Argumentverständnis, Argumentevaluation, Lernen, Vokabeln, Lesen, Schlussfolgern

Students who lack argument comprehension and evaluation skills will be at a significant disadvantage throughout their lives. These skills are used in many academic situations, including learning in science (Kuhn, 2010), and are necessary for many careers and professions. Argumentation skills are also beneficial for personal and civic decision-making. Unfortunately, many students finish high school with serious deficiencies in using argumentation skills for reasoning about science (e.g., Jimenez-Aleixandre, Rodrigues, & Duschl, 2000; Kuhn, 1993; Osborne, Erduran, & Simon, 2004) and general issues (Kuhn, 1991; Larson, Britt, & Kurby, 2009; Larson, Britt, & Larson, 2004).

Over the past two decades, there has been an increase in research on argumentation. Much of this work has investigated students' skill in producing arguments (Coquin-Viennot & Coirier, 1992; Golder & Coirier, 1994; Kuhn 1992; Knudson, 1992; 1994; Nussbaum, Kardash, & Graham, 2005; Stein & Miller, 1993). More recently, research has focused on whether there are beneficial effects of dialogue and peer interactions on middle school students' argumentation skills (Bell & Linn, 2000; Chinn & Anderson, 1998; Chinn, Duschl, Duncan, Buckland, & Pluta, 2008; Goldman, Duschl, Ellenbogen, Williams, & Tzou, 2003; Kuhn & Udell, 2003; Osborne, 2010; Reznitskaya et al., 2001), and efforts to use graphical tools to enhance argu-

DOI 10.1024/1010-0652/a000173

mentation skills (Bell, 1998; Dwyer, Hogan, & Stewart, 2012; Suthers, 2003). There has been, however, relatively little research on the basic cognitive processes involved in comprehending and evaluating simple arguments. Identifying these processes is critical for understanding more complex activities (e.g., writing persuasive prose or engaging in interactive arguments) and for predicting which pedagogical programs might improve students' performance on these activities.

Many studies of argumentation have based their analysis of the structure of informal arguments (see Argument 1a below) on the framework presented by Toulmin (1958). This framework describes a minimal informal argument as a claim and at least one supporting reason. The *claim* is a controversial statement that is not immediately accepted by the target of the discourse. The *reason* is a statement that supports the claim that is either acceptable itself or whose acceptability will be supported by another reason. For example, Argument 1a is an acceptable argument because it has a claim "Banks shouldn't charge ATM fees" and a reason "because paying fees makes customers unhappy." In contrast, Statement 1b is structurally flawed because it does not have a supporting reason.

1a. **Acceptable argument**. Banks shouldn't charge ATM fees because paying fees makes customers unhappy.
1b. **Unsupported claim**. Banks shouldn't charge ATM fees.
1c. **Unwarranted argument**. Banks shouldn't charge ATM fees because banks are financial institutions.

The other restriction of a minimal argument is that the reason must support the claim. Toulmin called the basis of this support the *warrant* – a general principle that makes the claim more likely to be true if the reason is true. Warrants are usually left unstated and it is up to the reader or listener to infer them and implicitly evaluate them. In Argument 1a, the warrant would be some principle like "Banks shouldn't do things that make their customers unhappy." Argument 1c is an example of an unwarranted argument because there is no readily inferable principle that connects banks being financial institutions to the claim that banks shouldn't charge ATM fees.

Recent studies (Britt, Kurby, Dandotkar, & Wolfe, 2008; Larson et al., 2009; Larson et al., 2004) have tested college students' ability to perform basic argument processing tasks, such as identifying the main claim in a short persuasive text or rejecting flawed items from among a series of two-sentence claim-reason pairs. They found that many college students have considerable difficulty with these basic tasks. For example, Larson et al. (2004) found that participants were only 30 % accurate at identifying the main claim and reasons from short, authentic, argumentative texts. Similarly et al. (2009) found that students frequently failed to reject poorly structured, simple arguments. These challenges in argumentation skills may be related to underdeveloped cognitive abilities in general (e.g., reading skill). By identifying the cognitive abilities that are most relevant to argumentation skills it may be possible to facilitate the development of these skills by providing training on these abilities. In the current work, we discuss argumentation skills in greater detail and present two studies that examined how these skills might be related to three general abilities: reading skill, reasoning skill, and vocabulary knowledge. These studies also assessed whether these general abilities influence the degree of improvement that students obtain from an online tutorial.

Cognitive argument processing skills

Argument comprehension and evaluation are fundamental to any persuasive discourse. In order to engage in debate or collaborative argument, one must, at a minimum, be able to understand and represent the opposing side's main claim and supporting reasons. Argument comprehension also underlies one's ability to evaluate one's own arguments and their likely reception by proponents of an opposing side.

Claim and reason detection and identification

To engage in argument processing, one must first detect that an argument is being put forth. Some claims may be detected in isolation as controversial statements that require support. Others may be recognized only indirectly as the statement for which support is given. Detection of reasons is similarly complex. To date, there is very little empirical evidence describing students' skill in claim and reason identification. One study by Chambliss (1995), using highly structured texts, found that skilled high school students were about 60 % accurate at stating the main claim when it was explicitly present and about 44 % accurate when the claim had to be inferred. Overall, students correctly identified only 79 % of the reasons, even though the reasons were presented in well-structured texts that elaborated each reason in its own paragraph and explicitly stated the claim-reason connection at the end of that paragraph. In more authentic texts, undergraduates show lower levels of claim and reason identification. Larson et al. (2004) found that college students were only 52 % ac-

curate at identifying explicitly stated main claims and 14 % accurate at identifying the reasons.

While we currently do not have a detailed theory of argument processing, the complex skill of argument comprehension requires the skill of representing the main point of a text. The identification of the main claim requires one to utilize lexical, syntactic, and semantic cues in the text. As with other genres, it is expected that students can use their knowledge of text cues and genre schemata to organize content (Blakemore, 1987, 2000; Meyer & Freedle, 1984; van Dijk & Kintsch, 1983). For argumentative texts, such genre signals for claims include evaluative terms (e.g., best, safest), modals (e.g., should) and qualifiers which can indicate either the generality of the claim (e.g., all, most, generally) or degree of confidence or certainty (e.g., maybe, possibly). Statements that include such modals and qualifiers are more likely to be judged as claims (Britt & Larson, 2003) and can affect the persuasiveness of the argument (Blankenship & Holtgraves, 2005; Durik, Britt, Reynolds, & Storey, 2008). Knowledge of such terms can be learned directly, as common in vocabulary learning, or it can be acquired indirectly through exposure to many expository texts. While the importance of the skill of abstracting the main point is clear for the main claim, it is also true of most reasons because reasons are often supported by elaborative or explanatory text.

Evaluating claim-reason support

A second essential argument skill is that of evaluating the degree of support that a reason provides. The informal arguments addressed in this research are reasons providing support for a particular claim. Unlike deductive arguments, the strength of a reason can only increase one's believability in the claim. One can evaluate arguments both in terms of their strength and structure. While there are many factors that will influence the degree of strength that a reader will assign to a particular reason, students must, at minimum, be able to determine when a reason is presented and whether that reason could provide support for the particular claim.

While there is a considerable body of research on how readers evaluate deductive arguments (e.g., Braine & O'Brien, 1998; Johnson-Laird, 1999; Johnson-Laird & Byrne, 1991; Rader, & Sloutsky, 2002; Rips, 1994) and on how argument strength and persuasiveness are related to attitude change (Lord, Ross & Lepper, 1979; Petty & Cacioppo, 1977; Schul & Goren, 1997), relatively few studies have examined individuals' evaluation of informal arguments. Larson et al. (2009) had undergraduates and 11th graders evaluate the structural quality of simple informal arguments. They gave participants 1 to 2 clause statements such as Arguments 1a-1c above and asked them to determine whether or not the statements formed a structurally acceptable argument without consideration of their own opinion and assuming the reason was true. The researchers found that the students had difficulty making this distinction. College students without training had accuracy rates of only 66 % (Experiment 1) and 62 % (Experiment 2), and high school students had an average accuracy rate of 70 %.

Britt et al. (2008) argued that many of these errors occur because readers process the claim superficially, especially the predicate (i.e., verb) such as "is immoral" in 2a or "is ineffective" in 2b (2a and b are shown below). In a series of experiments, they found that readers were not very accurate at recalling the predicate of the claim immediately after evaluating it. Instead, readers appeared to modify their representation toward a *gist* representation (e.g., "death penalty is bad") and lost the exact original wording. Such lack of precision can lead to reasoning errors in the context of evaluating support for an argument. For example, very different reasons support these two claims. Indeed, Britt et al. (2008) performed a tertiary split of their data on claim predicate recall and found that those that could accurately recall the claim predicate were more able to distinguish flawed arguments. Dandotkar et al. (in Press) replicated the role of precisely representing claim predicates when evaluating arguments with a new set of items.

2a. The death penalty is immoral.
2b. The death penalty is ineffective.

Tutoring evaluation skills

Fortunately, some of the difficulties students have with argument evaluation are easily remedied. Larson et al. (2009) presented students with a web-based tutor to enhance argument evaluation. To get the students' attention, the tutor first demonstrated to the students that they were lacking the skill and thus were vulnerable to manipulation (Koriat & Bjork, 2006; Sagarin, Cialdini, Rice, & Serna, 2002). Then the system tutored students to attend precisely to the claim predicate and how one determines whether a reason supports the claim. Interactive examples were followed by extended practice. The 35-minute tutor was found to be effective in improving argument evaluation skill. College students improved from 61 % to 76 % and high school students improved from 70 % to 88 % accuracy.

Despite the effectiveness of the tutor, there were still a substantial number of individuals who continued to make errors on these simple arguments. In fact, about half of the participants did not achieve mastery level of 80 % accuracy and about one-fifth of the participants were at or below

chance. This observation led us to speculate on which individual differences in pre-existing abilities might be related to argumentation skills. Determining the type of general cognitive abilities that are typically lacking in those having difficulty with each argumentation skill may help inform the design of future instructional interventions.

Individual differences in argument processing skills

Argumentation skills are clearly tied to other cognitive abilities. In order to comprehend a written argument, one must be a proficient reader. In order to properly evaluate argument quality, one must be a proficient reasoner. Based on theory and initial approaches to improving these skills, we have identified three general skill measures (vocabulary knowledge, reading ability, and analytic reasoning ability) that we expect will be related to argumentation skills.

Vocabulary knowledge

Although vocabulary knowledge is one component of reading ability and has been shown to be related to syllogistic reasoning skill (Siddiqui, West, & Stanovich, 1998; Stanovich, West, & Harrison, 1995) and inferencing skill (Osana, Lacroix, Tucker, Idan, & Jabbour, 2007), it may be important to treat acquisition of vocabulary as a separate general ability in these studies. There are two reasons for this. First, based on Britt et al.'s (2008) work indicating the importance of precision, we believe that general vocabulary measures might be a better indicator of how precisely a reader might encode a claim predicate than would general reading ability. Individuals who know more words might be more sensitive to fine distinctions of meaning and therefore more likely to preserve the verbatim representation of the claim. This skill is especially important since so much of the argument comprehension and evaluation tasks requires subtle distinctions among terms (e.g., "has a right" (factual), "is right to intervene" (evaluative), "should intervene" (policy)) and knowledge of specialized reasoning connectors (e.g., but, not, except, however).

The second reason for treating vocabulary separately is that expository texts often require knowledge not present in the text to make inferences (McNamara, Kintsch, Songer, & Kintsch, 1996) so one's word knowledge and its functional use might play a larger role in argument comprehension and evaluation. Readers who know more words would likely have better *lexical quality* associated with those words. Perfetti (2007) describes *lexical quality* as how the reader's knowledge of the word accurately represents the word's meaning and functional use. So, readers with higher *lexical quality* for words will likely know more contexts in which predicates are used and therefore more features of the thematic roles licensed by the predicates. Thematic roles are collections of constraints on objects that participate in the action described by a predicate. For example, in claim 3a, the predicate "should ban" has two thematic roles: an *agent* role, filled here by "the government", and a *patient* role filled by "cell phone use while driving."

3a. The government should ban cell phone use while driving.
3b. The government should ban cell phone use while driving because it is as dangerous as drunk driving.
3c. The government should ban cell phone use while driving because cell phones can be used to report accidents.

The predicate places constraints on the type of object that can fill each role. Specifically, not every object can fill the patient role of potentially being banned. For example, the prospective banned object must be harmful if the decrease or elimination of it would provide some benefit. We propose that the possible features of the thematic roles can determine the potential reasons that can support a particular claim predicate. For example, the reason in Argument 3b presents support that the patient is harmful and therefore can support the claim predicate. In contrast, the reason in 3c does not support the claim because it describes a cost rather than a benefit of banning the patient. We expect that individuals with higher vocabulary knowledge might be more aware of such thematic role constraints and thus better able to reject flawed arguments.

Reading ability

One would naturally expect that more proficient readers would perform better on written argument comprehension tasks. While no study has directly compared good and poor readers on argument comprehension skills, those same factors that lead to better comprehension of narrative and expository texts may also lead to better comprehension of argumentative texts. Skilled reading requires students to construct multiple levels of representation including a verbatim representation of the exact wording and structure of the statement as well as an inference-elaborated model of the situation described by the text (i.e., Situation Model) (Kintsch, 1988; van Dijk & Kintsch, 1983). Skilled readers are more likely to create an accurate and detailed verbatim representation of the actual text.

For example, Hannon and Daneman (2004) found that skilled readers were more likely to detect an anomaly (e.g., burying survivors) than less-skilled readers. They concluded that skilled readers create a more accurate representation whereas less-skilled readers construct a shallow or gist representation. In evaluating arguments, Britt et al. (2008) found that skilled readers also had more precise memory for both claim predicates and themes than less-skilled readers but the difference was even greater for claim predicates.

Skill differences are also found in the construction of a Situation Model, especially the important skill of elaborating the explicit text with inferences from within the text (i.e., bridging inferences) or using knowledge outside the text (e.g., elaborative inferences). Skilled readers are more likely to create a bridging inference (Cain, Oakhill, Barnes, & Bryant, 2001; Long, Oppy, & Seely, 1994) and less-skilled readers' comprehension performance increases with instructions to bridge (Hannon & Daneman, 1988). A second type of inference is the skill of determining the main point of a text. Skilled readers are better able to summarize a text (Winograd, 1984) and recall the most important information in a text (O'Reilly & McNamara, 2007). Because the main claim is the main point of an argument, this skill of abstracting the main point may be brought to bear in comprehending arguments, at least in identifying the main claim.

Analytic reasoning skill

In general, people not trained in logic have difficulty making valid conclusions, especially when their beliefs conflict with the necessary logical decision (e.g., Johnson-Laird, 1999; Stanovich et al., 1995). Although it is unclear whether such inferences are made based on syntactic rules (Braine & O'Brien, 1991; 1998) or semantics (Johnson-Laird & Byrne, 1991), the processes underlying one's skill in making these deductive inferences may be similar to warrant inferencing in argumentation. A critical aspect of both argumentation and deductive reasoning is the relationship between two or more elements. In the case of argumentation, the relationship is a warrant that licenses the claim from the presented reason or support. However, this relationship is probabilistic and domain-specific (Toulmin, 1958). In the case of deductive reasoning, if the premises are true and the structural relationship is valid, then the conclusion is definitely true. Because the skill of analyzing the relationship between components is so similar, we expected that deductive reasoning skill may help predict argumentation skill.

We selected the Analytic Reasoning Subtest of the LSAT (Law School Admission Test) as our measure of deductive reasoning. These problems assess analytic reasoning by presenting a scenario with arbitrary elements (e.g., factories labeled F, G, H, J, and Q) and their relationship with another variable (e.g., the days that a factory was inspected). The relationships generally involve grouping items or sequencing events. A short scenario is followed by a set of rules that require functional knowledge of logical relations (e.g., "or", "if", "and") and temporal relations (e.g., "before", "immediately after"). The reader is given a set of questions for each problem scenario that involves selecting a statement that either must be true/false or could be true/false. These items tap a reader's skill in representing and combining rules, and making valid deductive inferences.

Overview of studies

The following two studies examined factors that predict important argument skills. In Study 1 we examined the extent to which three general cognitive skills (vocabulary skill, reading comprehension skill, and analytic reasoning skill) predict a student's ability to comprehend and evaluate the quality of simple informal arguments. We also examined whether these cognitive skills can be used to identify the students that need additional support in learning to evaluate argument quality from a simple tutor. Study 2 examines more carefully a potential reason for the effect of one of these factors (vocabulary skill) on argument evaluation.

Study 1

In Study 1, U.S. undergraduates completed two argument comprehension tasks. Participants read a set of authentic arguments and identified the main claim (claim identification) and then underlined the reasons that supported the main claim they identified (reason identification). Then they completed the argument evaluation task in which they read a set of claim-reason arguments to decide whether the arguments were structurally acceptable or flawed without considering their own opinion. Then they completed a tutor designed to teach students how to evaluate arguments and performed the argument evaluation task again, with a new set of items. Finally, they completed three cognitive skills tests: vocabulary knowledge (Nelson-Denny), reading comprehension skill (section of the Law school admission test), and analytic reasoning skill (a section of the LSAT).

Research question 1 examined why some readers are better at simply identifying claims and reasons than oth-

ers. The task of claim identification requires that the reader abstract the main point of a text and state it at the correct level of precision. Skilled readers, compared with less-skilled readers, should be at an advantage because they create a more accurate representation (Hannon & Daneman, 2004). Readers with better vocabulary should also be able to make precise distinctions and represent texts at the optimal level of abstraction.

The task of reason identification requires that the reader encode each potential reason and verify the connection of the reason to the main claim. As with claim identification, it is expected that vocabulary skill will predict performance on the comprehension of reasons. The more knowledge one has of the possible thematic roles for a particular claim predicate, the better able one is to determine that a statement provides support for a particular claim. It is also possible that skilled analytic reasoners will also be more skilled at determining whether the claim is (at least partially) warranted from the reason. This claim-reason connection in informal arguments may involve constructing a generalized inference (i.e., warrant). Not all readers appear to verify identified reasons according to those that are warranted (Larson et al., 2004) but this step is critical to accurate reason identification. It may be the case that the skill of making deductive inferences is similar to the skill of inferring a warrant.

Research question 2 examined why some readers are better at evaluating simple arguments for quality than other readers. The argument evaluation task presented one-to-two clause statements (e.g., Arguments 1a–1c) and students determined whether the item was a structurally flawed argument. Again, precision is important to argument evaluation, especially precise encoding of the predicates of both the claim and the reason. Thus, one has to attend to subtle distinctions such as "has a right" (factual), "is right to intervene" (evaluative), "should intervene" (policy). Because knowledge of thematic roles for the claim predicate is important in making the acceptable/unwarranted distinction, vocabulary knowledge should predict skill in argument quality judgments. We also expect that analytic reasoning skill will be important because skill at creating logical inferences from deductive premises may play a role in creating warrant inferences. Such generalized inferences may be based on a skill similar to that of inferring the warrant of an informal argument.

Research question 3 examined why some readers learn from the argument evaluation tutorial whereas others do not. Prior research (Larson et al., 2009) demonstrated the effectiveness of a tutor to help students improve their evaluation skills; however, not all students benefited from the training. In an effort to improve the effectiveness of training argumentation skill for all students, we examined the characteristics of the students that fail to achieve mastery after completing the training. Because the tutor teaches students to attend to the claim predicate and test the reason against that claim predicate, it is expected that those with lower vocabulary knowledge and who are less skilled in analytic reasoning should have more difficulty learning this skill from a short explanation and admonitions to be precise. Given the tutorial was written, it is also expected that less skilled readers should have more difficulty learning from this type of tutor. Therefore, it is expected that all three skills may be important in discriminating skilled from less skilled learners.

Method

Participants

One hundred sixty-one native English speaking undergraduate psychology students from Northern Illinois University participated in this experiment for course credit.

Design and Procedure

Participants experienced each session in a private, sound attenuated room. All participants received the tasks in the same order in this two-session study. During session 1, participants completed the set of pre-test argument skill assessments (argument comprehension and pre-flawed-judgment task) and then completed the web-based computer tutor designed to help students develop the skills necessary for comprehending and evaluating arguments. The first session lasted approximately 1.25 hours. Participants returned two days later to complete the posttest assessment (post-flawed-judgment task) and the measures of cognitive ability (vocabulary skill, reading comprehension, and reasoning skill). To reduce the impact of fatigue on performance on any of these measures, the cognitive skill battery was presented in different orders using a latin square.

Argument Skill Assessments

Argument Comprehension

The ability to comprehend written essays containing arguments was evaluated by participants' skills in identifying the main claim and supporting reasons of authentic arguments. Participants were presented five written argument essays, based on Larson et al. (2004), and were asked to state the main claim and underline each supporting reason. Each argument had one main claim and between 3 and 8 reasons ($M = 5.20$, $SD = 1.92$). Prior to being given the packet, participants completed a practice argument to ensure they understood the task. This task took approximately 20 minutes to complete.

Claims were scored as correct if the participant mentioned both the predicate and the theme. For example, the

claim for one argument was "TV has a negative influence on children and family life." To receive credit, the participant would need to mention the theme (TV) and the predicate (has negative influence). Synonyms were not allowed because the article was present for the assessment. Two independent raters trained in the coding procedure achieved interrater reliability for the scoring of 20 % of the arguments with a kappa of .989 for claims. All discrepancies were discussed until agreement and one rater coded the remaining items.

Reasons were scored as correct if both the predicate and the theme were underlined as a single unit. Two independent raters trained in the coding procedure achieved interrater reliability for the scoring of 20 % of the arguments with a kappa of .991 for reasons. All discrepancies were discussed until agreement and one rater coded the remaining items. The identification of reasons was treated as a signal detection task because the task required one to differentiate the reason (i.e., a signal) from the noise (i.e., any other information not deemed a reason). Otherwise a student who simply underlined most of the text would get a very high score for correct identifications ("Hits") without correcting for all the incorrectly identified information ("False Alarms"). Hits and false alarms can be used to calculate measures of sensitivity and bias in signal detection analyses (e.g., Stanislaw & Todorov, 1999). Sensitivity reflects the ability to discriminate when a signal is present from when the signal was absent (i.e., the degree of overlap between the signal and the noise). Bias represents the general tendency to respond either "yes" or "no" during the task. Higher scores indicate greater sensitivity to the signal. Raw scores were converted to D prime scores for each individual.

Table 1. Flawed Judgment Task (FJT) Example Items.

Flawed Unwarranted Arguments
1. The money universities spend on athletics is a waste of campus resources because the biggest advertiser during sporting events is beer companies.
2. Handguns benefit society because handguns are protected by the second amendment of the constitution.
Flawed Unsupported Claim Arguments
3. Banning cell phone use while driving is unfair.
4. Natural foods are not always healthier than processed foods.
Acceptable Arguments
5. People should be allowed to have only two biological children because it would help stabilize population growth.
6. Americans should support a two-party political system because more than two parties will result in candidates being elected without a majority vote.

Flawed Judgment Task (FJT)

A 32-item evaluation task assessed the reader's ability to evaluate simple arguments (Dandotkar et al., in Press; Larson et al., 2009) with three types of arguments: acceptable arguments, unsupported claims, and unwarranted arguments (see 1a–1c for another example and discussion of these items). The same theme is presented here to illustrate how the items varied in each condition; however, participants received a different theme for each item. Themes were randomly assigned within condition.

5a. **Acceptable argument**. All children should be encouraged to engage in after school activities because they help keep children off the streets and involved in safe, regulated pastimes.

5b. **Unsupported claim**. All children should be encouraged to engage in after school activities.

5c. **Unwarranted argument**. All children should be encouraged to engage in after school activities because all children are small.

Participants were instructed to read each argument and identify structural flaws ignoring their own opinion and the perceived strength of the argument. So 5a should have been identified as structurally acceptable and 5b and 5c should have been identified as structurally flawed (see Table 1 for additional example items from the FJT task). Of the 32 items, 16 were good arguments and 16 were flawed (8 unwarranted and 8 unsupported claim). The length of the arguments did differ among types. As expected, the unsupported claims contained fewer words on average (M = 10.6, SD = 3.5) than the unwarranted (M = 17.13, SD = 5.8) and the acceptable arguments (M = 20.56, SD = 5.8). If readers use a length heuristic, this should make improvement even easier. There were two versions of the test so that one could be presented as a baseline measure (pre-FJT) and one could be presented after the tutor to assess improvement (post-FJT). These two versions were counterbalanced across participants. This task took approximately 10–15 minutes to complete. Both versions of the FJT were scored for the number of correct judgments. The two versions of the FJT had acceptable internal consistency (Cronbach's alpha = .813 and .795).

Argument Tutorial

To improve students' argument evaluation skills, we used the web-based version of the Larson et al. (2009) tutor. The tutor began by motivating students to learn the new skill. Students first evaluated a set of 6 flawed arguments. If they accepted any of the arguments, the tutor revealed that they were fooled and warned that they are vulnerable to being tricked by others through non-logical means. Then the module defined and gave examples of well-structured and flawed arguments followed by steps for identifying each type. Stu-

dents were warned about two common problems that arise when trying to distinguish flawed from structurally acceptable arguments (failing to disregard one's own opinion of the truth of the claim and being overly rejecting). Throughout the tutorial, students were tested on their understanding of the material. Finally, students received a set of practice items. For each item, they first clicked on the claim predicate. This step is important because less skilled reasoners do not consistently attend to the claim predicate, and those who can recall the predicate after evaluating the argument are better skilled at evaluating the quality of the argument (Britt et al., 2008). Then, students evaluated the argument by selecting "yes" for structurally acceptable arguments and "no" for flawed arguments. Students received immediate explanatory feedback. Then they completed another set of practice trials in which they had to make the evaluation judgment from memory and type the claim predicate into a textbox. This recall was automatically scored using an LSA-based matching algorithm and feedback was given immediately. Our past work found that skilled reasoners often kept the predicate active in memory even after reading the entire argument whereas less-skilled reasoners often did not (Kurby, Britt, & Dandotkar, 2006, July). Because memory of the claim is necessary for making a warranted-unwarranted judgment, students need practice holding the predicate in memory while reading the reasons. Interaction with this tutorial took approximately 35 minutes.

Measures of cognitive ability

Vocabulary Assessment

Vocabulary knowledge was assessed by a shortened version of the Nelson-Denny vocabulary test (Form F; Brown, Fishco, & Hanna, 1993). In this test, participants must choose the synonym of the word from a short list of options. Because the test becomes progressively more difficult, we created a modified version by selecting every other item to reduce the test to 50 items. Responses were coded as either correct or incorrect and a proportion score was computed for each participant. The reliability for this vocabulary measure is quite high (split-half method = .90; Cunningham & Stanovich, 1997). Participants were given 7.5 minutes to answer as many of the items as they could.

Reading Comprehension Assessment

Two of our measures of general abilities (reading comprehension and reasoning ability) come from the Law School Admissions Test (LSAT). Performance on the combined test correlates with academic successes in the first year of law school (Anthony, Harris, & Pashley, 1999) and the subtests each have re-test reliability coefficients between .90 to .95 (Law School Admission Council, 2001). The Reading Comprehension portion of the LSAT was used as a measure of the participant's ability to critically read expository texts. In this comprehension assessment, participants read expository passages (approximately 400 words) on topics such as physical and social sciences, law, art, and the humanities. Each passage was followed by several questions requiring the participant to verify a paraphrase, make appropriate inferences, identify the structure of the passage, or reason about the material in the passage. An abridged version of a test was used which took fifteen minutes to complete. Two passages were presented to each participant followed by six questions for each passage. Responses were coded as either correct or incorrect and a proportion score was computed for each participant.

Analytic Reasoning Assessment

In this test of deductive reasoning from the Analytic Reasoning portion of the LSAT, participants are presented a scenario and set of rules (see Table 2) organized around group membership or sequencing. The solver has to validly combine rules and then use deduction to select an acceptable sequence or grouping from that set of rules (see question 1 in Table 2) or in a situation with an additional hypothetical rule (see question 2). For each scenario, participants are given 4 to 6 questions that require an accurate understanding of terms critical for making deductive inferences such as "if", "or", "not", "must", "except." Many

Table 2. Example scenario and questions from the LSAT Analytic Reasoning Assessment.

During a period of six consecutive days – day 1 through day 6 – each of exactly six factories – F, G, H, J, Q, and R – will be inspected. During this period, each of the factories will be inspected exactly once, one factory per day. The schedule for the inspections must conform to the following conditions:

F is inspected on either day 1 or day 6.

If G is inspected on day 3, Q is inspected on day 5.

J is inspected on an earlier day than Q is inspected.

Q is inspected on the day immediately before R is inspected.

1. The inspection scheduled for day 3 and day 5, respectively, *could* be those of
A) G and H
B) G and R
C) H and G
D) R and J
E) R and H

2. If the inspection of R is scheduled for the day immediately before the inspection of F, which one of the following must be true about the schedule?
A) The inspection of either G or H is scheduled for day 1.
B) The inspection of either G or J is scheduled for day 1.
C) The inspection of either G or J is scheduled for day 2.
D) The inspection of either H or J is scheduled for day 3.
E) The inspection of either H or J is scheduled for day 4.

questions also require the combining of several rules simultaneously. For the purposes of this study, participants were given two scenarios each followed by 5 questions. Participants were allowed 15 minutes to complete this task. Responses were coded as either correct or incorrect and a proportion score was computed for each participant. The subtests each have re-test reliability coefficients between .90 to .95 (Law School Admission Council, 2001).

Measures of interest

Interest in the task was measured both before and after the tutor in order to consider motivational as well as cognitive factors in predicting argumentation skill. The five interest items (e.g., "Analyzing arguments is interesting" and "This activity is boring," reversed) were focused on the positive feelings associated with the task. Each item was rated from 1 (*Strongly disagree*) to 7 (*Strongly agree*). The scale showed high internal consistency as measured by Cronbach's alpha both pre-tutor (.86) and post-tutor (.88).

Results

Preliminary analyses and Overview of Primary Results

First, we examined version and order effects. Recall that different versions of the FJT were used for pretest and posttest, and the order of these versions was counterbalanced across participants. A MANOVA compared scores on the FJT across versions and neither multivariate nor univariate effects emerged as significant. Moreover, participants completed the cognitive ability measures in one of three random orders. A MANOVA comparing these three orders on the cognitive ability measures also revealed no significant effects.

We organized the primary results into four sections. First, using multiple regression analysis, we examined the extent to which the cognitive ability measures predicted pre-tutor argumentation skills (i.e., claim and reason identification, initial FJT performance). Second, we compared participants' task performance pre-tutor and post-tutor using paired samples t-tests. Third, returning to multiple regression, we used the cognitive ability measures to predict pre-tutor performance on the FJT as well as post-tutor performance, controlling for pre-tutor performance.

The means, standard deviations, and correlations for all variables are presented in Table 3. One person had missing data on three variables, which explains the slight variation in degrees of freedom across analyses. Two patterns can be observed in the matrix. First, although the cognitive ability measures are positively correlated with each other (range from .12 to .27), they do not seem to be so highly related to each other as to warrant a concern about redundancy or multicolinearity. Second, because neither pre-interest nor post-interest positively predicted argumentation skill, these variables were excluded from the primary analyses.

Table 3. Correlations, means, and standard deviations for all variables in Study 1.

		1	2	3	4	5	6	7	8	9
1.	FJT-pre									
2.	FJT-post	.54**								
3.	Claim ident.	.40**	.51**							
4.	Reason ident.	.41*	.49**	.60**						
5.	Vocabulary	.30*	.36**	.22**	.23**					
6.	Analytic reasoning	.19	.27**	.10	.23**	.17*				
7.	Reading comprehension	.13	.21**	.21**	.12	.27**	.12			
8.	Pre-interest	−.06	−.04	−.09	<.01	.05	−.02	.02		
9.	Post-interest	−.08	−.06	−.17*	−.01	−.04	.07	.01	.73**	
	M	.64	.74	.55	1.30	.46	.29	.25	4.68	4.81
	SD	.15	.16	.03	.50	.01	.02	.13	1.15	1.24
	N	161	161	160	160	161	161	161	161	160

Notes. Performance variables are represented in proportion of items correct, with the exception of reason identification, which is calculated as d-prime. Interest could range from 1 (*low interest*) to 7 (*high interest*). FJT = Flawed Judgment Task.
*$p < .05$. **$p < .01$

Cognitive Abilities and Initial Argumentation Skills

Claim identification

The mean percent accuracy for claim identification is presented in Table 3. The three cognitive ability variables (vocabulary, analytic reasoning, and reading comprehension) were entered simultaneously on the same step of the regression to predict initial (pre-tutor) argument comprehension skills (claim identification and reason identification) and argument evaluation skills (FJT). Both R^2 and sr^2 are reported as measures of effect size for multivariate and univariate effects, respectively. In addition, standardized regression coefficients are reported in the text, and unstandardized coefficients and standard errors are reported in the tables.

The cognitive ability variables accounted for a significant amount of variance in claim identification overall, $F(3, 157) = 4.23$, $p < .01$, $R^2 = .08$. Two significant univariate effects emerged. Vocabulary positively predicted claim identification, such that participants with more advanced vocabulary skills were more able to identify claims than those with less advanced vocabulary skills, $t(157) = 2.09$, $p = .04$, $\beta = .17$, $sr^2 = .03$. Reading comprehension also positively predicted participants' ability to identify argument claims, $t(157) = 1.98$, $p = .05$, $\beta = .16$, $sr^2 = .02$. Analytic reasoning did not uniquely predict claim identification (see Table 4).

Reason identification

The average D prime score for reason identification is presented in Table 3. The three cognitive ability measures were entered on the same step of the regression to predict reason identification. The cognitive ability variables accounted for significant variability overall, $F(3, 156) = 5.31$, $p < .01$, $R^2 = .09$. Two univariate effects emerged (see Table 4). Vocabulary positively predicted participants' performance, $t(156) = 2.36$, $p = .02$, $\beta = .24$, $sr^2 = .03$. Those with higher versus lower vocabulary scores were more able to correctly identify the reasons provided to support arguments. Analytic reasoning positively predicted reason identification, $t(156) = 2.43$, $p = .02$, $\beta = .19$, $sr^2 = .03$, but reading comprehension did not.

Initial FJT performance

When predicting FJT performance prior to the tutor, the three cognitive ability variables predicted a significant portion of variance when entered together, $F(3, 157) = 6.62$, $p < .01$, $R^2 = .11$. One univariate effect emerged. Participants with higher vocabulary scores performed better than those with lower vocabulary scores, $t(157) = 3.41$, $p < .01$, $\beta = .27$, $sr^2 = .07$. Whereas vocabulary positively predicted argument evaluation, neither analytic reasoning nor reading comprehension were significant predictors (see Table 4).

Comparing Measures Pre- and Post-tutor

The next analysis compared participants' scores on the FJT across time, from pre-tutor to post-tutor. A paired sample t-test revealed that participants' performance improved from before ($M = .64$, $SD = 0.15$) to after ($M = .75$, $SD = 0.16$) the tutor, $t(160) = -9.05$, $p < .01$. In other words, participants learned from the tutor and/or practice with these items.

Cognitive Abilities and Improvement from Pre- to Post-tutor

Multiple regression was used to examine the extent to which participants' improvement on the FJT from pre- to post-tutor could be predicted by the candidate cognitive variables. This regression was performed in two steps. First, pre-tutor performance was entered on the first step, yielding a significant positive effect, $t(159) = 7.99$, $p < .01$, $\beta = .54$, $R^2 = .29$, $sr^2 = .29$. The three cognitive ability measures were added on the second step, and contributed a significant increase in variability accounted for in post-tutor performance, $F(3, 156) = 5.72$, $p < .01$, $\Delta R^2 = .07$. Pre-tu-

Table 4. Results of regression analyses predicting argument comprehension and evaluation for Study 1.

	Claim identification		Reason identification		Pre-tutor FJT		Post-tutor FJT	
Predictor	B	SE	B	SE	B	SE	B	SE
Pre-tutor FJT	--	--	--	--	--	--	.48*	.07
Vocabulary	.34*	.16	.56*	.24	.24*	.07	.17*	.07
Analytic reasoning	.10	.14	.52*	.21	.12	.06	.13*	.06
Reading comprehension	.40*	.20	.19	.30	.04	.09	.11	.08

Notes. The regression coefficients are unstandardized. FJT = Flawed Judgment Task.
*$p < .05$

tor performance, $t(156) = 6.51$, $p < .01$, $\beta = .44$, $sr^2 = .17$, remained significant on this step. Both vocabulary score, $t(156) = 2.51$, $p = .01$, $\beta = .18$, $sr^2 = .03$, and analytic reasoning, $t(156) = 2.17$, $p = .03$, $\beta = .14$, $sr^2 = .02$, positively predicted improvement from pre- to post-tutor. However, reading comprehension was not a significant predictor of change, $t(156) = 1.33$, $p = .19$, $\beta = .09$, $sr^2 = .01$ (see Table 4).

Discussion

Study 1 showed that both readers with high vocabulary knowledge and better reading skill were more able to identify main claims in authentic arguments while those with high vocabulary knowledge and better reasoning skill were more able to identify reasons in those arguments (research question 1). We also found that readers with high vocabulary knowledge were better able to evaluate simple arguments for quality (research question 2). Finally, we found that readers with high vocabulary knowledge and better reasoning skill were more able to learn from the tutorial to evaluate simple arguments (research question 3). The most consistent finding was the importance of vocabulary knowledge in argument comprehension and evaluation while controlling for other skills, like reading comprehension and analytic reasoning.

Study 1 does not, however, distinguish between the *breadth* (how many words one knows) and *depth* (how much meaning is connected to the words one knows) of the participants' vocabulary knowledge (Nagy & Herman, 1987). It may be that participants who know more words also have a deeper functional understanding of those words which allows them to better comprehend and evaluate the arguments that they read. This distinction is important for designing a more effective intervention.

Study 2

The intention of the second experiment was to focus on the contribution of vocabulary knowledge on the ability to acquire new argumentation skills. One central component of acquiring this skill is the ability to assess the relationship between the claim and its reason (the warrant). One important element of this process may be the thematic role that is activated by the predicate of the claim. Thematic roles are the semantic relations for predicates. Two key roles are the agent (e.g., monster in 6a or baby in 6b), the person or thing doing the action, and the patient (e.g., baby in 6a and monster in 6b [6a and b are shown below]). McRae, Ferretti, and Amyote (1997) wanted to determine the features of the agent and patient roles for a set of predicates. They had participants list the features of the agent ("someone who frightens people") and the patient ("someone who is frightened"). They found that people did indeed list common features. For example, common features of the agent of frighten were mean, scary, ugly, and big, whereas the common features of the patient of frighten were scared, small, weak, and helpless. These features help explain why the particular nouns used to fit the thematic roles in 6a seem to fit better than in 6b. The characteristics of a monster overlap better with the common features of the agent of frighten than the characteristics of a baby. In a reading time study, Ferretti, McRae, and Hatherell (2001) found that when reading the predicate (e.g., arresting), the features of thematic roles (agent-cop, patient-criminal) were immediately activated.

6a. The monster frightened the baby.
6b. The baby frightened the monster.

Features of thematic roles should also be available during claim processing and therefore may guide the determination of a reason-claim relationship. Based on Ferretti et al. (2001), Britt, Larson, Millis, & Jordan (July 2008) used a feature listing task to elicit the thematic roles of argument predicates. Undergraduates were given verbs for policy claims (e.g., *discourage, approve*, *ban, mandate*) and verbs of evaluation (*detrimental, beneficial, valuable, worthwhile*) and were asked to list features of the agent (e.g., "Someone who bans things is ...") and the patient (e.g., "Something that is banned is ..."). Across the set of experiments, they found that participants agree upon a set of key features, those features fit a meaningful and consistent pattern, the set of features accurately captured the set of reasons supporting claims in arguments from a corpus analysis, and that the features predict strength and acceptability ratings.

In Study 2, we examined whether specific knowledge about features of thematic roles for argument predicates would predict skill on the FJT, and whether knowledge of the thematic roles would emerge as plausible mediating variables between vocabulary skill and FJT performance. Thus, participants completed the FJT and the general vocabulary test from Study 1. They also completed a feature listing task which required attention only to the patient which is most important for reason selection. It was predicted that vocabulary knowledge would be positively correlated with argumentation evaluation skill and that knowledge about features of thematic roles (feature listing) would overlap with this relationship. This is because the knowledge of features is more specifically relevant to determining whether a reason can support a claim than is knowledge of synonyms.

Method

Participants

Sixty-four native English speaking undergraduate psychology students from Northern Illinois University participated in this experiment for course credit.

Materials

To investigate our question of interest, eight verbs were chosen that depict persuasive behaviors by individuals. These verbs could influence how the thematic role of the claim predicate would be perceived. The critical verbs were: prohibit, support, allow, discourage, require, permit, ban, and encourage.

Flawed Judgment Task (FJT)

This argument evaluation task was similar to the one used during Study 1 with a few alterations. This version of the FJT contained 48 items containing the three types of arguments mentioned above (good arguments, unwarranted arguments, and unsupported claims). In addition to the original 32 items, 16 news items were included to assess evaluation skills when these target verbs were present. Half of the 48 items were acceptable arguments and half were flawed (12 unwarranted and 12 unsupported claims). Each of the target verbs was present in one good argument and in one flawed argument, so each participant experienced each of these verbs two times. Similar to the materials of Study 1, 48 different themes were randomly assigned to each argument. Participants were never exposed to more than one version of a single theme. This version of the FJT had acceptable internal consistency (Cronbach's alpha = .732). The instructions were the same as in Study 1. The time it took to complete this task ranged from 15 to 20 minutes.

Vocabulary Assessment

An enhanced version of the Nelson-Denny vocabulary assessment used in Study 1 was used. Eight items were added to this assessment to make sure that any problems with vocabulary were not due to individuals' lack of knowledge of the critical verbs (Cronbach's alpha = .88). These additional eight items were used to measure individuals' specific vocabulary scores whereas the remaining items were used to measure the individuals' general vocabulary scores. Participants were allowed nine minutes for completion of this task.

Predicate Schema Listing

Participants were required to list features associated with particular verbs. A list of 16 verbs were presented using a computer program hosted by E-prime. Eight of these verbs were the critical verbs mentioned above and eight were filler items. The task of the participant was to list features of the thing that is acted upon. So when a verb was presented, participants were asked to produce the features of the thing as implied by the verb. For example, features of something that is thrown may be that it is small, round, hard, light, etc. Note that participants were not asked to produce the actual things that are thrown (i.e. a ball, a rock, a frisbee). The program presented one verb at a time and participants had two minutes to list as many features as they could think of. Features were categorized into 1 of 28 categories. The categories represented descriptive qualities of the predicate such as "Caring", Fair", "Acceptable", "Obedient", "Entertaining", "Beneficial", "Deadly." Twenty percent of the responses were scored by two independent scorers trained in the procedure with an agreement of 92 %. Disagreements were discussed until agreement and one rater coded the remaining data.

Design and Procedure

Study 2 incorporated a within subject design in which all participants were exposed to the same materials and tasks in the following order: predicate schema listing, FJT, vocabulary assessment. The predicate schema listing and the vocabulary assessment were timed tasks as indicated above. The FJT was not timed and participants could take as long as they needed to complete it. Each participant experienced each session in a private, sound attenuated room. The predicate schema listing was presented using a personal computer. The FJT and the vocabulary assessment were presented in the paper-and-pencil fashion. This one-session study lasted approximately 60 minutes.

Results and Discussion

Correlations, means, and standard deviations for all of the variables in this study are presented in Table 5. The primary aim of Study 2 was to test a mechanism through which vocabulary might predict skill in evaluating arguments. Specifically, it was hypothesized that the relationship between vocabulary score and FJT performance on target words may be accounted for by participants' ability to generate the appropriate descriptors for objects associated with argument predicates (predicate schema listing). To test this hypothesis, a series of analyses were conducted consistent with Baron and Kenny's (1986) approach to testing mediation (see Table 6). First, we tested the extent to which general vocabulary predicted performance on the FJT. Similar to Study 1, participants' general vocabulary score was a significant and positive predictor, $t(62) = 2.88$, $p < .01$, $\beta = .34$, $sr^2 = .12$. In a second regression analysis, we tested the extent to which general vocabulary also predicted participants' performance on the predicate listing task.

Table 5. Correlations, means, and standard deviations for all variables in Study 2 (N = 64).

		1	2	3	4
1.	FJT				
2.	Predicate schema listing	.41**			
3.	General vocabulary	.34**	.37**		
4.	Specific vocabulary	.34**	.30*	.43**	
	M	.66	18.61	.50	.80
	SD	.15	9.43	.17	.13

Notes. Flawed judgment task (FJT) and vocabulary scores equal the proportion of items correct. Predicate schema listing represents the total number of characteristics generated.
*$p < .05$. **$p < .01$

This also revealed a significant effect, $t(62) = 3.14$, $p < .01$, $\beta = .37$, $sr^2 = .14$, thereby providing evidence of the second requirement for mediation. Finally, both general vocabulary and predicate schema listing were used to predict performance on the FJT. This analysis showed that predicate schema listing was a positive predictor of argument evaluation, $t(61) = 2.69$, $p < .01$, $\beta = .33$, $sr^2 = .09$, and that the effect of general vocabulary was reduced to marginal significance, $t(61) = 1.81$, $p = .08$, $\beta = .22$, $sr^2 = .04$. A Sobel test revealed that a statistically significant portion of variance in the relationship between general vocabulary and argument evaluation could be accounted for by the predicate schema listing scores, $z = 1.99$, $p < .05$. A test of the indirect effect using a bootstrapping approach showed the same result, 95 % CI (0.02, 0.24). These analyses provide converging evidence that the predicate schema listing is a plausible mediator of the effect between general vocabulary and argument evaluation. See Figure 1 for a depiction of these relationships.

Finally, it was important to test whether the relationship between vocabulary and predicate schema listing is related to general vocabulary ability or to participants' understanding and knowledge of the particular words used in the predicate schema listing task. To examine this, both general and specific vocabulary scores were used simultaneously to predict predicate schema listing. The results of this analysis showed that only general vocabulary, $t(61) = 2.30$, $p = .03$, $\beta = .30$, $sr^2 = .07$, and not specific vocabulary, $t(61) = 1.32$, $p = .19$, $\beta = .17$, $sr^2 = .02$ predicted predicate schema listing. Although it is worth noting that the general vocabulary score is a standardized test whereas the specific vocabulary score is not, the specific vocabulary score is matched to the actual task it is predicting. The significant effect of general vocabulary may suggest that the precision with which individuals used words in general and not knowledge of these particular words contributed to the effect on argument evaluation.

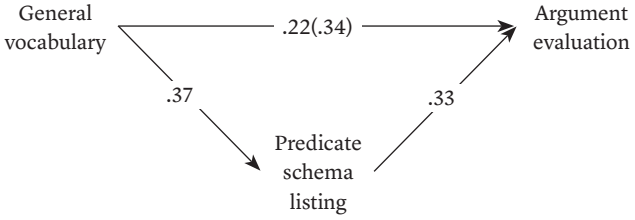

Figure 1. Mediation model tested in Study 2. This is a path diagram depicting the relationship between general vocabulary and argument evaluation, as mediated by predicate schema listing. The numbers are standardized regression coefficients. The number in parentheses represents the relationship between general vocabulary and argument evaluation when predicate schema listing was not in the analysis.

General Discussion

Across two studies we found that all three cognitive abilities positively predicted argument comprehension (i.e., reading skill and vocabulary knowledge predicted claim identification while reasoning skill and vocabulary knowledge predicted reason identification) and evaluation performance (i.e., vocabulary knowledge predicted argument evaluation skill). The most important finding, however, was the importance of vocabulary knowledge. General vocabulary knowledge was a better predictor of performance on the argument evaluation task than was specific knowl-

Table 6. Results of regression analyses testing mediation in Study 2.

	Criterion variable					
	Argument evaluation				Predicate schema listing	
Predictor	B	SE	B	SE	B	SE
General vocabulary	.31*	.11	.20	.11	21.15*	6.73
Predicate schema listing	--	--	.005*	.002	--	--

Notes. The regression coefficients are unstandardized.
*$p < .05$

edge of the predicates used. Thus, the effect was not simply the result of knowing the words used in the test, but more likely something about general knowledge for words. Finally, as expected, we found a relationship between all three cognitive abilities and improvement from the tutorial. Vocabulary knowledge was the strongest predictor and was significant when controlling for reading and reasoning skill (Study 1).

In Study 2, we found that this vocabulary-evaluation relationship was mediated by individuals' functional knowledge of the predicates. When evaluating arguments, understanding exactly what is meant by a predicate is extremely important because it restricts the set of reasons that can support a given claim. Those with less precise knowledge of predicates are less able to determine if the provided reason supports the claim, and therefore are also less able to correctly evaluate arguments.

The importance of vocabulary for argument evaluation is consistent with the Lexical Quality hypothesis (Perfetti & Hart, 2002; Perfetti, 2007), which emphasizes the role of readers' knowledge of and skill in integrating events described across sentences. Just like the current findings, Stafura and Perfetti (2014) found that vocabulary knowledge was a better predictor of word-to-text integration processes than were other measures of reading skill. For argumentation, integration of information across clauses and sentences is exactly what is needed to infer a claim-reason warrant and subsequently evaluate this relationship. Thus, both argument comprehension and evaluation require such processing.

Limitations

Although the results suggest that targeting functional vocabulary (predicate schema) should help students who do not learn from a simple tutorial, we have not yet tested this directly. We also have not yet looked for any possible relationship between performance on the predicate schema listing task and argument comprehension. Although we would expect that predicate schema listing might predict claim identification, it may turn out to be a better predictor of reason identification.

One caveat and potential area for future investigation is that the claim predicates for the FJT were all either policy claims (arguing for a behavior) or value claims (arguing for a belief about an evaluation or judgment of something). We need much more work to better understand the evaluation of factual (arguing that something existed, exists, or will exist) or causal claims (arguing for one or more causal relationships for something).

Our materials required general knowledge rather than domain-specific knowledge. However, domain-specific argumentation is not necessarily very different. Certainly, the general skill of evaluating whether there is a reason and whether the relationship is minimally connected would hold for several types of domain specific arguments. Likewise, readers commonly encounter policy and evaluative claims in discipline-specific discourse: science (e.g., "We should reduce carbon emissions"), social science ("Watching violent TV programs is harmful; It is inappropriate for the United States to intervene in other countries' affairs") and history ("The launch of Sputnik I was a significant event for the U.S."). The need to attend to claim predicates and evaluate predicate schema would therefore be the same for domain-specific argumentation as it is for more general reasoning. Differences would likely come with factual and causal claims, where domain-specific knowledge provides additional constraints on the acceptability of reasons beyond those of the predicates themselves. These types of differences might appear more readily in a strength rating task rather than in a minimal quality evaluation task. With domain-specific epistemic standards for the best evidence or support, training within a discipline would lead to variation in how much weight should be given to a particular support.

One final limitation is that the correlations, while significant and consistent, indicate that there are likely other factors that, if identified, could increase the effectiveness of argument tutorials. We were specifically targeting malleable factors that can be trained rather than personality characteristics. It is likely that characteristics such as need for cognition or skepticism may predict one's base level of skill or ability to learn from the tutorial. Such personality characteristics, while not malleable, may be useful in tailoring training approaches to specific types of students or styles of learning.

One factor not considered here that has recently been shown to not effect performance is an over reliance of less-skilled reasoners on low-level word overlap between argument elements (Dandotkar et al., in Press). If anything, conceptual or word overlap actually helped skilled reasoners accept good arguments but did not hurt less-skilled reasoners. Another, more concerning factor is the tendency of less-skilled reasoners to accept warrants or claims based on personal agreement (Lord et al., 1979). In fact, a post hoc analysis of the Dandotkar et al. (in Press) data found that less-skilled reasoners were more likely to accept attitude-consistent flawed arguments than were skilled reasoners (see Dandotkar, 2012). Britt et al. (2008) also found that less-skilled readers' memory for claims was more affected by their personal agreement with the claims than skilled readers. Thus, another general area for instruction might be to encourage evaluation of arguments without reference to one's own opinion. In our study, the tutor warned students about the need to do this but a warn-

ing may not be sufficient. Less-skilled reasoners and less-skilled readers may need direct practice on using this skill.

Educational implications and future directions

Current models of argument processing do not explain how one determines that a reason supports a claim, which makes it challenging to design tutorials. The current studies are a small step in this direction (see also Voss, Fincher-Kiefer, Wiley, & Silfies, 1993). First, we found that comprehension and evaluation skill were significantly correlated (r = .41 to .60). Further work should more closely examine the progression of these skills within an individual student. This will allow us to better understand the natural progression of such skills. For example, it may be easier to teach students to comprehend the main claim and this would undoubtedly be a precursor to comprehending and evaluating reason quality.

These results point to a method for enhancing the effectiveness of our tutorial for those students most in need. Given that reading skill predicted improvement from the tutorial, perhaps it should be revised to be written at a simpler reading level. The most important result, however, is that those with the greatest improvement from the current tutorial were those that had the most vocabulary knowledge (Study 1) and knowledge of predicate schemas in particular. Participants less familiar with specific argument predicates may not have a well-developed predicate schema. Thus, students could be taught schemas for classes of claim predicates and they could use them to guide the identification of argument elements, to systematically find weaknesses in an argument, to verify that an argument is warranted and to construct stronger, more complete, arguments. This is not to say that such information alone would be sufficient. Students may have this knowledge but may not understand its relevance for argument evaluation so they would still need to understand the specific criteria for argument quality. Studies that combine tutorial interventions with controlled materials help us to understand students' abilities and how to improve their reading capabilities.

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education in part by Grant R305F100007 to UIC and Grant R305H05133 to Northern Illinois University and by the Center for the Interdisciplinary Study of Language and Literacy (CISLL) at Northern Illinois University. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences, the U.S. Department of Education or CISLL.

References

Anthony, L. C., Harris, V. F., & Pashley, P. J. (1999). *Predictive validity of the LSAT: A national summary of the 1995–1996 correlation studies 3.* Law School Admission Council research report series technical report, 97-01. Netwon, PA: Law School Admission Council (LSAC).

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Bell, P. L. (1998). *Designing for students' science learning using argumentation and classroom debate* (Unpublished doctoral dissertation). University of California at Berkeley, Berkeley, CA.

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education, 22*, 797–817.

Blakemore, D. (1987). Linguistic constraints on pragmatic interpretation: A reassessment of linguistic semantics. *Behavioral and Brain Sciences, 10*, 712–713.

Blakemore, D. (2000). Indicators and procedures: Nevertheless and but. *Journal of Linguistics, 36*, 463–486.

Blankenship, K. L., & Holtgraves, T. (2005). The role of different markers of linguistic powerlessness in persuasion. *Journal of Language and Social Psychology*, *24*, 3–24.

Braine, M. D., & O'Brien, D. P. (1991). A theory of if: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98*, 182–203.

Braine, M. D., & O'Brien, D. P. (Eds.). (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

Britt, M. A., Kurby, C. A., Dandotkar, S., & Wolfe, C. R. (2008). I agreed with what? Memory for simple argument claims. *Discourse Processes, 45*, 52–84.

Britt, M. A., & Larson, A. (2003). Construction of argument representations during on-line reading. *Journal of Memory and Language, 48*, 749–810.

Britt, M. A., Larson, A. A., Millis, K. K., & Jordan, J. P. (2008, July). *Features of claim predicates help determine whether a reason supports a claim*. Paper presented at the annual meeting of the Society for Text and Discourse, Memphis, TN.

Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denny Reading Test: Manual for scoring and interpretation*. Itasca, IL: Riverside.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition, 29*, 850–859.

Chambliss, M. J. (1995). Text cues and strategies successful readers use to construct the gist of lengthy written arguments. *Reading Research Quarterly, 30*, 778–807.

Chinn, C. A., & Anderson, R. C. (1998). The structure of discussions that promote reasoning. *Teachers College Record, 100*, 315–368.

Chinn, A. C., Duschl, R. A., Duncan, R. G., Buckland, L. A., & Pluta, W. J. (2008). A microgenetic classroom study of learning to reason scientifically through modeling and argumentation. In G. Kanselaar, J. van Merrienboer, P. Kirschner, & T. de Jong (Eds.), *International perspectives in the learning sciences: Creating a learning world: Proceedings of the Eight International Conference for the Learning Sciences* (pp. 14–15). Utrecht, The Netherlands.

Coquin-Viennot, D., & Coirier, P. (1992). Structures of argumentative discourse: Effects of type of referential space. *European Journal of Psychology of Education, 7*, 219–229.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*, 934–945.

Dandotkar, S. (2012). *Readers' reliance on semantic and logical relatedness when evaluating arguments.* Doctoral dissertation, Northern Illinois University.

Dandotkar, S., Magliano, J. P., & Britt, M. A. (in press). The effect of logical relatedness and semantic overlap on argument evaluation. *Discourse Processes*.

Durik, A. M., Britt, M. A., Reynolds, R., & Storey, J. K. (2008). The effects of hedges in persuasive arguments: A nuanced analysis of language. *Journal of Language and Social Psychology, 27*, 217–234.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2012). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition and Learning, 7*, 219–244.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language, 44*, 516–547.

Golder, C., & Coirier, P. (1994). Argumentative text writing: Developmental trends. *Discourse Processes, 18*, 187–210.

Goldman, S. R., Duschl, R. A., Ellenbogen, K., Williams, S., & Tzou, C. T. (2004). Science inquiry in a digital age: Possibilities for making thinking visible. In H. van Oostendorp (Ed.), *Cognition in a digital age* (pp. 253–283). Mahwah, NJ: Erlbaum.

Hannon, B., & Daneman, M. (1988). Facilitating knowledge-based inferences in less-skilled readers. *Contemporary Educational Psychology, 23,* 149–172.

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

Jimenez-Aleixandre, M., Rodrigues, M., & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education, 84*, 757–792.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology*, *50*, 109–135.

Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction*. Hove, UK: Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension construction-integration model. *Psychological Review, 95*, 163–182.

Knudson, Ruth E. (1992). The development of written argumentation: An analysis and comparison of argumentative writing at four grade levels. *Child Study Journal, 22*, 167–184.

Knudson, Ruth E. (1994). An analysis of persuasive discourse: Learning how to take a stand. *Discourse Processes, 18*, 211–230.

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory & Cognition, 34*, 959–972.

Kuhn, D. (1991). *The skills of argument*. New York, NY: Cambridge University Press.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*, 155–179.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77*, 319–337.

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education, 94*, 810–824.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*, 1245–1260.

Kurby, C. A, Britt, M. A., & Dandotkar, S. (2006, July). *Representing argument claims: Availability and accessibility of the predicate and implications on argument evaluation*. Poster presented for the 16th Annual Conference of the Society for Text and Discourse, Minneapolis, MN.

Larson, A. A., Britt, M. A., & Kurby, C. (2009). Improving students' evaluation of informal arguments. *Journal of Experimental Education, 77*, 339–365.

Larson, M., Britt, M. A., & Larson, A. (2004). Disfluencies in comprehending argumentative texts. *Reading Psychology, 25*, 205–224.

Law School Admission Council. (2001). *10 actual, official LSAT preptests*. Newton, PA: Law School Admission Council, Inc.

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1456–1470.

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*, 2098–2109.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition & Instruction, 14*, 1–43.

McRae, K. T., Ferretti, T., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes, 12*, 137–176.

Meyer, B. J., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal, 21*, 121–143.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. C. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19–35). Hillsdale, NJ: Erlbaum.

Nussbaum, E. M., Kardash, C. M., & Graham, S. E. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology, 97*, 157–169.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43*, 121–152.

Osana, H. P., Lacroix, G. L., Tucker, B. J., Idan, E., & Jabbour, G. W. (2007). The impact of print exposure quality and inference construction on syllogistic reasoning. *Journal of Educational Psychology, 99*, 888–902.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*, 463–466.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in science classrooms. *Journal of Research in Science Teaching, 41*, 994–1020.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357–383.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Washington, DC: American Psychological Association.

Petty, R. E., & Cacioppo, J. T. (1977). Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and Social Psychology, 35*, 645–655.

Rader, A. W., & Sloutsky, V. M. (2002). Processing of logically valid and logically invalid conditional inferences in discourse comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 59–68.

Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32*, 155–175.

Rips, L. J. (1994). *Psychology of Proof.* Cambridge, MA: MIT Press.

Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. (2002). Dispelling the illusion of invulnerability: The motivations and

mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology, 83*, 526–541.
Schul, Y., & Goren, H. (1997). When strong evidence has less impact than weak evidence: Bias, adjustment, and instructions to ignore. *Social Cognition, 15*, 133–155.
Siddiqui, S., West, R. F., & Stanovich, K. E. (1998). The influence of print exposure on syllogistic reasoning and knowledge of mental-state verbs. *Scientific Studies of Reading, 2*, 81–96.
Stafura, J. Z., & Perfetti, C. A. (2014). Word-to-text integration: Message level and lexical level influences in ERPs. *Neuropsychologia, 64*, 41–53.
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavioral Research Methods, Instruments, & Computers, 31* (1), 137–149.
Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology, 31*, 811–826.
Stein, N. L., & Miller, C. A. (1993). The development of memory and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 285–335). Hillsdale, NJ: Erlbaum.
Suthers, D. D. (2003). Representational guidance for collaborative inquiry. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environment* (pp. 27–46). Netherlands: Springer-Verlag.
Toulmin, S. E. (1958). *The uses of argument*. Cambridge, MA: Cambridge University Press.
Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
Voss, J. F., Fincher-Kiefer, R., Wiley, J., & Silfies, L. N. (1993). On the processing of arguments. *Argumentation, 7*, 165–181.
Winograd, P. N. (1984). Strategic difficulties in summarizing texts. *Reading Research Quarterly, 19,* 404–425.

Manuscript submitted: 17.3.2015
Manuscript accepted: 20.9.2016
Conflict of interest: No

Anne Britt
Northern Illinois University
Dekalb, IL, 60115
USA
britt@niu.edu