Analyzing Argumentation In Rich, Natural Contexts

ALINA REZNITSKAYA   *Montclair State University*

RICHARD C. ANDERSON   *Center for the Study of Reading University of Illinois at Urbana-Champaign*

Abstract: The paper presents the theoretical and methodological aspects of research on the development of argumentation in elementary school children. It presents a theoretical framework detailing psychological mechanisms responsible for the acquisition and transfer of argumentative discourse and demonstrates several applications of the framework, described in sufficient detail to guide future empirical investigations of oral, written, individual, or group argumentation performance. Software programs capable of facilitating data analysis are identified and their uses illustrated. The analytic schemes can be used to analyze large amounts of verbal data with reasonable precision and efficiency. The conclusion addresses more generally the challenges for and possibilities of empirical study of the development of argumentation.

Résumé: On présente les aspects théoriques et méthodologiques de la recherche sur l'apprentissage de l'argumentation des enfants d'école élémentaire; décrit un encadrement théorique des mécanismes psychologiques responsables pour cet apprentissage et pour le transfert des discours argumentatifs; illustre quelques applications de cet encadrement avec suffisamment de détails pour guider les futures investigations empiriques des exécutions argumentatives orales, écrites, individuelles et en groupe. On identifie des logiciels capables de faciliter l'analyse de données, et on illustre leurs applications; les schèmes analytiques peuvent s'employer avec précision et efficacité pour analyser des grandes quantités de données verbales. L'article termine avec une description générale des possibilités des études empiriques sur l'apprentissage de l'argumentation et des défis que celles-ci auront à relever.

Keywords: argumentation, discourse analysis, argument schema, cognitive processes, psychological theory, social learning, performance-based assessment, measurement, research methodology, mixed methods, computer applications, technology, idea unit, elementary schools, persuasive writing

1. Introduction

Many contemporary educators maintain that students should learn to comprehend, construct, and evaluate reasoned arguments in order to competently deal with the complexities of their professional and personal lives (e.g., Kuhn, 1992; Lipman, 1991; Voss & Means, 1991). Cultivating argumentation ability is seen as essential

for enabling students "not just to think, but to think well" (Kuhn, 1991, p. 1).

The study of argumentation development must rely on theoretically-driven analytic approaches that allow us to capture, represent, and interpret important features of argumentative discourse. In recent years, there have been impressive advances in methodology designed to examine classroom interactions and related educational outcomes (e.g., Applebee, Langer, Nystrand, & Gamoran, 2003; Chinn & Anderson, 1998; Means & Voss, 1996; Nystrand, Wu, Garmon, Zeiser, & Long, 2003; L. Resnick, Salmon, Seitz, Wathen, & Holowchack, 1993). However, in studies focused on substantive, rather than on methodological issues, the level of detail in descriptions of analytic procedures is often insufficient to provide clear directions to researchers interested in conducting their own analyses of students' communications (e.g., Applebee et al., 2003; Means & Voss, 1996). At the same time, studies primarily focused on methodology often present analytic schemes that are so complex and labor intensive that they could hardly be applied to a large corpus of data (e.g., Chinn & Anderson, 1998; L. Resnick et al., 1993). For example, Resnick et al. (1993) used a fine-grained analytic framework to examine two minutes of a group discussion, and Chinn and Anderson (1998) presented an analysis of only nineteen discussion turns. Neither study elaborated on the ways to automate the proposed methodology, or to otherwise make it suitable for larger datasets. Further, in both substantive and methodological studies, little or no attention is given to the use of technology for enhancing and facilitating data-analytic procedures (e.g., Applebee et al., 2003; Chinn & Anderson, 1998; Means & Voss, 1996; L. Resnick et al., 1993). In rare instances where an applicable software program is mentioned (e.g., Applebee et al., 2003), there is little or no description of exactly how it was used to organize and interpret the data.

In this paper, we address existing shortcomings in the literature on researching argumentation by describing specific strategies used to examine a large corpus of complex, multifaceted verbal data. Importantly, many methodological strategies described below may be applied to investigation of any group or individual performance involving argumentation. We will first discuss pedagogical models and theoretical propositions that motivate our research, including *Collaborative Reasoning* model, *Argument Schema Theory*, and the *Snowball Hypothesis*. Next, we will illustrate selected analytic procedures and related software programs used to generate concise numerical summaries of rich, qualitative-type data. Finally, we will address more broadly the challenges and opportunities involved in analyzing naturally-occurring arguments.

2. Pedagogical and Theoretical Frameworks

Research described in this paper aims to address an important educational goal of helping students become skilled in the discourse of reasoned argumentation. Reasoned discourse allows claims to be critically examined, it respects the laws of evidence and logical principles, and it is based on the rules that can be known,

studied, and practiced. This is the discourse used to resolve a variety of important issues ranging from scientific controversies to guilt or innocence in murder trials. Sadly, it is also the discourse that is rarely present in a typical American classroom (Almasi, O'Flahavan, & Arya, 2001; Alveraman, O'Brien, & Dillon, 1990; Nystrand et al., 2003). Thus, it comes to no surprise that numerous nation-wide assessments and research studies consistently document the lack of proficiency in argumentation by the majority of American students (e.g., Kuhn, 1991; McCann, 1989; Means & Voss, 1996; NAEP, 1994, 1999, 2002).

Collaborative Reasoning (CR) is an instructional approach that attempts to acquaint elementary school children with argumentative discourse (Waggoner, Chinn, Yi, & Anderson, 1995). Typically, during CR discussions, students gather in small groups to discuss a central question from the story they have read. Stories are selected to contain moral, social, or scientific dilemmas that are engaging for young children and can stimulate a thoughtful dialog. For example, one of the CR stories, *Amy's Goose* (Holmes, 1977), is about a lonely farm girl who befriends a goose that has been injured by a fox. Amy wants to keep the goose as a pet instead of letting it fly south with the rest of the flock. The discussion question is, "Should Amy let the goose go?", and the story contains evidence that can support contrasting resolutions of this issue.

Every CR discussion is different depending on the story, the composition of the group, or the amount of disagreement among the students. However, CR discussions share several common elements. During the discussions, children are expected to take a public position on the issue, support it with reasons and evidence from the story, and challenge other discussion participants with counterarguments and rebuttals. Students in CR discussions decide when to talk and what to discuss. The teacher' role is to provide scaffolding for the development of argumentation and student management of turn taking. Importantly, the emphasis in CR discussions is not on reaching a consensus on the issue. Rather, we want students to experience the process of rational judgment. The ultimate goal of CR includes "inculcating the values and habits of mind to use reasoned discourse as means for choosing among competing ideas" (Anderson, Chinn, Waggoner, & Nguyen, 1998, p. 172).

CR model is derived from an explicit theoretical framework, called *Argument Schema Theory* (AST) (Reznitskaya & Anderson, 2002). AST integrates independent research traditions, including 1) argumentation theory developed by philosophers (e.g., Govier, 1985; Toulmin, 1958; van Eemeren, Grootendorst, & Henkenmans, 1996; Walton, 1996), 2) schema-theoretic views of cognition (Anderson, 1977; Chambliss & Murphy, 2002; Meyer, Brandt, & Bluth, 1980; van Dijk & Kintsch, 1983), and 3) the study of social influences on learning (L. Resnick et al., 1993; Rogoff, 1990; Vygotsky, 1962; Wertsch, 1985).

Following normative models of a rational argument (e.g., Toulmin, 1958), we postulate the elements of *an argument schema,* or an abstract knowledge structure that can be instantiated with context-specific details. A developed argument schema

includes such elements as the statement of belief, reasons, grounds, warrants, backing, modifiers, counterarguments, and rebuttals. It contains the understanding of the rhetorical organization of an argument, its properties, functions, and conditions for use. While different domains (i.e., moral, scientific, legal) may have their own argumentation standards (Toulmin, 1958), even these "field-dependent" rules can be generalized across multiple contexts. Thus, we can think of argumentative knowledge as an aggregation of field-invariant and field-dependent rules, principles, and informal heuristics, which together comprise an argument schema. Importantly, an argument schema is more than a simple collection of individual elements. Rather, the elements and their relationships are supported through a set of epistemological beliefs, which constitute an "explanatory framework" (Mishra & Brewer, 2003) for the schema. An *evaluative* type of epistemology (Kuhn, 1991) represents the normative structure. The evaluative view assumes that knowledge is situated in a given context, while also recognizing that some judgments are more reasonable than others (cf., Hofer & Pintrich, 1997; King & Kitchener, 1994; Perry, 1970)

Research on schematic structures identified important influences of a schema on perception, comprehension, learning, inferencing, and remembering (e.g., Anderson & Pichert, 1978; Brewer & Treyens, 1981; Chambliss, 1995; Cheng & Holyoak, 1985; Meyer et al., 1980). Generalizing from this research, we hypothesize the functions of a developed argument schema to include: (1) allocating attention to argument-relevant information; (2) directing retrieval of argument-relevant information from memory and permitting inferential reconstruction; (3) organizing argument-relevant information; (4) providing the basis for anticipating objections and for finding flaws in one's own arguments and the arguments of others; and (5) facilitating argument comprehension, construction, and repair (Reznitskaya & Anderson, 2002).

To explain the acquisition of an argument schema we draw upon social theories of learning (e.g., Luria, 1981; Mead, 1962; Rogoff, 1990; Vygotsky, 1962; Wertsch, 1985). These theories emphasize the priority "in time and in fact" of social interaction in individual learning. Participation in social settings allows children to observe, try out, and eventually internalize various "psychological tools" (Vygotsky, 1981) that advance their cognitive development to higher levels. The educational potential of social activity comes from its dialogic organization (Bakhtin, 1981, 1986; Kuhn, 1992; Mead, 1962; Vygotsky, 1981). De-emphasizing the distinction between public argument and private thinking, Bakhtin writes that "our thought itself…is born and shaped in interaction and struggle with other's thought, and this cannot but be reflected in the forms that verbally express our thoughts as well" (Bakhtin, 1986, p. 92). Similarly, Mead views individual reasoning as a process of internal argumentation, a dialog with a "generalized other" (Mead, 1962, p. 156). The ability to incorporate the voices of "others" into one's own thinking comes from engagement in social settings, where participants collectively formulate, defend, and scrutinize multiple viewpoints.

By integrating social learning theories with schema-theoretic views of cognition, we are able to further specify psychological mechanisms underlying the development of reasoning. Specifically, Anderson et al (2001) propose that an argument schema can be broken down into recurrent verbal patterns, or *argument stratagems.* Argument stratagems are rhetorical and reasoning moves utilized in argumentation. They serve a variety of cognitive and social functions. For example, children in CR discussions often find it beneficial to appeal to their previously read story for evidence. Using such expressions as "in the story it said" or "on page 23 she said", they explicitly mark information as coming from the story in order to enhance its credibility and add to the persuasive force of their argument. We labeled this stratagem with the general form "In the story it said [EVIDENCE]". The capitalized, bracketed part of the stratagem will change in response to contextually different scenarios. However, the underlying purpose, form, possible consequences and objections to this stratagem will remain the same. During group discussions, children pick up and reuse effective argument stratagems they see other children using, an idea referred to as the *Snowball Hypothesis* (Anderson et al., 2001). According to the Snowball Hypothesis, useful stratagems spread among children and tend to occur in discussions with increasing frequency.

An important question related to the Snowball Hypothesis is whether the increased use of argument stratagems represents simple mimicry. That is, have students internalized the deeper meaning of a stratagem or are they merely parroting word strings? This question can be addressed through careful examination of the immediate conversational context of the stratagem use, evaluating whether the conditions are appropriate. Another indicator of the mindful usage is the ability to flexibly change the surface form of the stratagem, while preserving its discourse function.

The Snowball Hypothesis provides an empirically researchable account of argumentation development *during* oral group discussions. In order to further examine the degree to which individual students internalized argumentative knowledge, we need to step out of the social context. Will engagement in group discussions help students perform better on argument-related tasks when social support is no longer available? We propose that abstract properties of a schema should enable transfer of argumentative knowledge. Just like entering a new restaurant activates a "restaurant schema" (Schank & Abelson, 1977) abstracted from multiple prior experiences with eating out, an encounter with a task requiring the use of argumentation should trigger a set of cognitive and social practices that constitute an argument schema. For example, given an individual task involving the use of argumentation, such as a written persuasive composition, students should rely on the argument schema to generate, organize, and edit the content. That is, they should focus on proposing reasons for the taken position, anticipating counterarguments, and offering rebuttals. They can be expected to properly utilize argument stratagems, such as "In the story if said, [EVIDENCE]," generalizing knowledge acquired through participating in CR discussions to the new argument-

related task, performed individually. Although it has been argued that a switch from oral to written argumentation and from group to individual performance may reduce the possibility of transfer (Freedman & Pringle, 1988; Pellegrini, Galda, & Rubin, 1984), there is emerging evidence that such generalization is possible, especially when the schema is sufficiently developed (Dong, Anderson, Li, & Kim, 2006; Kuhn, Shaw, & Felton, 1997; Reznitskaya, Anderson, & Kuo, 2007; Reznitskaya et al., 2001).

Similarly, students with the developed argument schema should interact differently with an argumentative text. Once the text is recognized as an argument, readers should proceed to make use of the 'slots' in the activated schema, looking for claims, supporting reasons, counterarguments, and rebuttals.

To summarize, in this section we discussed a specific pedagogical model, CR, and articulated related psychological theory. According to AST and the Snowball Hypothesis, argument schemas are developed in group settings, where children pick up and use functional argument stratagems introduced by innovative group members. Once internalized, the knowledge of argumentation can be transferred to new situations, allowing students to perform better on individual tasks, such as writing a persuasive composition or reading an argumentative text. We will now turn to empirical evaluations of the theoretical principles just described and the related data analytic approaches.

3. Empirical Studies and Data-Analytic Strategies

Analyzing Group Argumentation

Anderson and his colleagues (2001) conducted a study to empirically test the *Snowball Hypothesis*, or to investigate the ways in which children acquire and subsequently reuse various argument stratagems. Fourth grade students in this study engaged in a total of forty-eight CR discussions, which were videotaped and transcribed by researchers. Sifting through transcripts of these discussions, we tracked the occurrence of thirteen argument stratagems. These speech acts served various functions, including managing participation, positioning oneself in relation to a classmate's argument, acknowledging uncertainty, extending the story world, using story information as evidence, etc. For example, in order to gain the floor for a classmate, children frequently used the stratagem of the general form "What do you think, [NAME]". Children were also able to modify the surface form of the stratagem, while maintaining its discourse function. For instance, a phrase "Would you like to share anything?" was also used by the children to manage group participation.

We wanted to examine the possibility of diffusion or contagion of identified argument stratagems from a single child to others in a group. Models of diffusion through social network have been previously investigated in a variety of contexts (Bohstedt, 1994; Lefebvre, Whittle, Lascaris, & Finkelstein, 1997; Morris, 1994),

although their applications to education are surprisingly rare (cf., Roth, 1996). We relied on several methodological strategies used by previous researchers in order to systematically evaluate the hypothesized social mechanism of argumentation development. The first strategy involves calculating transition probability. This is the probability that the event (E) will happen again, given that it has occurred a certain number of times already. The symbolic expression for transition probability is $P(E+1|E)$. For example, we calculated the likelihood of an argument stratagem "In the story it said [EVIDENCE]," to be used the third time, given that it has already occurred twice in previous discussions. In addition, we calculated the median number of lines in discussion transcripts before each new occurrence of the stratagem, counting from its previous use.

We also compared two probability models suitable for the analysis of categorical data: Random Poisson and Contagious Poisson. The frequency distribution of independent, random events is represented by the Random Poisson distribution. The Contagious Poisson distribution is another model that has an additional parameter indicating the extent to which prior events (i.e., first occurrence of an argument stratagem in a discussion) change the likelihood of subsequent events (i.e., additional occurrences of the same stratagem). We analyzed observed distribution of thirteen argument stratagems in relation to expected distributions under both models, identifying which model better represents the data.

Before we could apply the mathematical procedures just described, we needed to account for every instance of a particular argument stratagem appropriately used by children in forty-eight CR discussions. This required an analysis of an enormous corpus of data, consisting of 14,942 lines of children's naturally occurring discourse. We will now turn to the detailed description of specific steps taken to prepare the data for mathematical analyses. As the goal of this paper is to share with future researchers of argumentation a variety of data-analytic procedures suitable for large datasets, computer programs that greatly facilitated the analysis will also be described in detail.

TransTool (Kumar & Miller, in press) is a computer software helpful for processing digital video. With TransTool, one can view, transcribe, timestamp, and code video files. Transcripts with related codes can then be exported into Word, Excel, or other programs. The main advantage of transcribing with TransTool is that it rapidly cues the video to the point marked with the timestamp or an assigned code, allowing for an easy access to a given scene. This feature speeds up the transcription process and enhances the accuracy of resulting records. In addition, it facilitates re-examination of the immediate context of a proposition, including the non-verbal aspects of interactions.

Because TransTool has very limited data-analytic functions, QSR NUD*IST (QSR, 1997) was used to further examine discussion transcripts. QSR NUD*IST is a qualitative data analysis software that allows for flexible searching, coding, and analyzing text patterns. Consider, for example, an analysis of one of the CR

discussions, based on the *Amy's Goose* story (Holmes, 1977). During the CR discussion, a student who believes that Amy should let the goose go, makes the following comment: "But in the story it said that he was well enough to go and fly". This is an appropriate use of the argument stratagem "In the story it said [EVIDENCE]. In order to identify *all* uses of this stratagem, we systematically searched each discussion transcript using QSR NUD*IST *Text Search* function and its variations (regular, approximation, etc). Below we present selected results of a QSR NUD*IST search for the word string 'in the story':

…But later on <u>in the story</u>, she says, it says that she thinks that, um, he goose really is strong enough, that he can go. She just doesn't want to let him go, because she likes him.

But <u>in the story</u> it said that he was well enough to go and fly.

Cause, it, um, the gander would probably die too, because, um, <u>in the stor</u>y it says, when they were flying away, all, when they were all far away, all of a sudden (alone) the goose (pulled) back and (sees) the gander, and it was like, and um, the gander had come back many times to the, um, barn uh, calling for his mate…

We searched for alternative word strings that conventionally would have the same or nearly the same meaning as the most typical expression of each stratagem. We tried to cast a wide net by thinking of different ways to express ideas ourselves and by remaining alert to children's modes of expression. When another wording variation was discovered, we reran a search to make it more inclusive. For example, we ran additional Text Searches each time we encountered a new surface form of a stratagem "In the story it said [EVIDENCE]", including such statements as "On page 32, she said [EVIDENCE]", "The story tells you [EVIDENCE]", etc. QSR NUD*IST made it easy to examine alternative surface forms of structurally equivalent stratagems, helping to generate evidence regarding the extent to which peer modeling leads to individual internalization.

Next, the results of the Text Search were coded using QSR NUD*IST *Free Nodes*. Free Nodes allow for storing information about a particular text segment. For example, to code the results of the Text Search just discussed, we highlighted the identified uses of the stratagem and marked them with a node, called "Story Evidence".

In the following step, every speech act provisionally coded with Free Nodes was evaluated by two raters working independently. QSR NUD*IST program permits viewing coded information in an enclosing paragraph or a larger section. The raters' task was to evaluate the discourse function of each utterance. Re-

examination of all identified stratagems by multiple raters, made feasible using QSR NUD*IST, helped to reduce the inherent subjectivity in judgment. Our raters agreed on classification 97% of the time. Such high degree of agreement lends credibility to the data-analytic procedures, as well as to the ensuing conclusions.

Because we used QSR NUD*IST software we were able to conduct a fine-grained analysis of an enormous amount of data, identifying, coding, and reviewing 1,631 instances of thirteen argument stratagems used by the children. The feature of the program that allows a coded word string to be effortlessly placed back into its conversational context made it possible to examine not only the linguistic form of a stratagem, but also its function, meaning, and condition of use. This, in turn, permitted the required contextual sensitivity, which is often absent when natural discourse is fragmented into easily quantifiable segments. In a recent study of group dynamics, Li et al. (2007) used TransTool in combination with QSR NUD*IST to review the coded moves on videotape, thus getting an even more nuanced contextual interpretation.

Once we were satisfied with the reliability of the coding scheme, we used QSR NUD*IST *Profile* feature in order to summarize coded information in terms of the number of characters, words, lines, etc. The generated summaries were exported into Microsoft Excel. Using the latter program, we calculated the transition probability for each argument stratagem and the median number of lines in discussion transcripts before a given occurrence. Table 1 displays the results of this analysis for the previously discussed stratagem, "In the story it said [EVIDENCE]" across all 48 CR discussions. The table shows that there were more lines before the first occurrence than before second, and more lines before second than before later occurrences. Similar pattern was found for 12 other argument stratagems identified through the analysis, thus providing support to the Snowball Hypothesis.

Table 1
Likelihood And Spacing Of 'In The Story, It Said [EVIDENCE]'

Measure	First	Second	Third	Fourth	Fifth
P(E+1|E)	0.79	0.87	0.76	0.64	0.84
Lines before	47	33	19	19	14

Also, using the data generated with the help of QSR NUD*IST, we were able to model the diffusion of argument stratagems, as shown in Table 2. The first column in Table 2 represents selected numbers of occurrences of any of the 13 argument stratagems in a discussion. The second column represents the observed frequency for a given number of occurrences. The next two columns represent modeled frequencies.

Table 2

Observed And Expected Frequency Of Argument Stratagems

Number of Occurrences	Observed Frequency	Modeled Frequency	
		Contagious Poisson	Random Poisson
0	311	263.6	46.5
1	55	101.3	120.7
2	61	62.6	156.7
3	47	43.6	135.7
4	19	32.0	88.1
5	22	24.3	45.8
6	18	18.8	19.8
7	23	14.8	7.4
…	…	…	…
…	…	…	…
…	…	…	…
28	0	0.2	0.0
29	1	0.2	0.0
Chi-Square		**9.78**	**2.91E+16**

From Table 2, the contagious Poisson distribution provides a better fit for the data (x^2=9.78, df=33, P>.5) than the random Poisson (x^2=2.91E+16, df =34, P<.01). This finding is, again, consistent with the Snowball Hypothesi*s*, indicating that the occurrence of an effective argument stratagem increases the likelihood of its later use.

The methodology for examining the diffusion of argument stratagems was applied in two recent studies of argumentation. In the first study, Kim et al. (in submission) looked at the spread of argument stratagems during 20 CR discussions conducted online, via web forums. The second study examined the occurrence of the stratagems in 24 CR discussions of children from a Chinese industrial city, a Chinese village, and a Korean city (Dong et al., 2006). In both studies, contagious Poisson model provided much better fit for the data than independent Poisson. These studies replicated the original findings from Anderson et al. study (2001). They extended the application of the Snowball Hypothesis to new communicative modes, as well as to cultural and linguistic contexts. Thus, credible generalizations about argumentation can emerge when tools permit systematic analysis of large quantities of carefully-collected naturalistic data.

Empirical evaluations of a Snowball Hypothesis illuminated psychological mechanisms that promote the acquisition of argumentative discourse during *group* interactions. The next question to address is whether or not engagement in group argumentation helps students to perform better on argument-related tasks performed individually.

Analyzing Individual Argumentation

AST suggests that abstract properties of knowledge structures acquired from enriching experience with argumentation should enable the flexible use of these structures in different contexts and communicative modes. A quasi-experimental study was conducted to investigate the transfer potential of social interactions (Reznitskaya et al., 2007). 128 elementary school children and their teachers participated in one of the three treatment conditions for a period of 5 weeks. Two classrooms were assigned to each condition. In the first treatment condition, students engaged in CR discussions and received explicit instruction in abstract principles of argumentation. The second treatment group participated in CR discussions and had no explicit instruction in argumentation. Finally, the third group received their regular reading instruction and did not have either CR discussion or explicit instruction. Students from all six classrooms completed two transfer tasks designed to measure their ability to apply the knowledge of argumentation acquired during the intervention to new situations. The tasks included 1) writing a persuasive essay and 2) recalling an argumentative text.

When delivering explicit instruction to the elementary school students in the first treatment condition, we employed the metaphor of building an argument being similar to building a solid house. Figure 1 (p. 186) can be conceptualized as a basic argument schema, which contains five parts of an argument, including position, reasons, supporting facts, objections, and responses to objections. Figure 1 also depicts the relationship among the argument components and exemplifies linguistic markers commonly used to introduce each component. For example, an objection is being introduced with "Some people might say".

The basic argument schema depicted in the figure is modeled after the "pyramid heuristic" Yeh (1998) used to teach argumentative writing to middle-school students. This formulation also borrows from the useful framework proposed by Toulmin, who pioneered the effort to identify nonoverlapping functions of argument components, including claims, grounds, warrants, backing, modifiers, and rebuttals (Toulmin, 1958; Toulmin, Rieke, & Janik, 1979). Notably, Toulmin's model does not explicitly include counterarguments. Following other scholars who consider opposing perspectives to be an important part of argumentation (Kuhn, 1991; van Eemeren & Grootendorst, 1984; Walton, 1996), we expanded the model to incorporate counterarguments, or objections. In contrast with the Toulmin model, the basic argument schema in our study omitted warrants, backing, and modifiers. The latter argument components were outside the scope of our investigation, which

Figure 1. A Basic Argument Schema

focused on the most crucial discourse elements that were already present in the arguments of young children or that could be introduced through developmentally-appropriate instruction. We used the basic argument schema to support the development of argumentative knowledge during oral discussions, as well as to evaluate student subsequent performance on two individual transfer tasks.

For the first task, students were asked to write a persuasive essay in response to a story similar to those used as a basis for the CR discussions. Briefly, in the story, a boy named Thomas wins the school Pinewood Derby race, but he breaks

the rules by not making his model car by himself. He confides to his classmate, Jack, that he has received help in making his car. Jack is faced with the dilemma of whether or not he should tell on Thomas. Students in three treatment conditions were asked to discuss Jack's dilemma in their essays.

Similar to the previously discussed analysis of oral group interactions, we used a qualitative data analysis software, in conjunction with other programs, in order to facilitate the discovery of interesting regularities related to the individual construction of written argument. The essays were analyzed using QSR Nvivo software (QSR, 1999). Developed by the same company as the previously discussed NUD*IST, QSR Nvivo is a similar program, which was chosen because its improved version became available at the time of the study.

The analysis of persuasive essays involved six steps. First, all essays were transcribed and given an anonymous identification number (ID) to keep researchers blind to treatment differences when evaluating students' responses. Next, we imported essays into QSR NVivo software as separate documents. In the following step, each document was parsed into *idea units*. An idea unit, as defined by Mayer (1985), "expresses one action or event or state, and generally corresponds to a single verb clause" (p. 71). More detailed rules for chunking student essays into idea units are discussed elsewhere (Reznitskaya et al., 2007; Reznitskaya et al., 2001).

In step four, we used *Free Nodes* to assign a unique code to each idea unit. For example, different free nodes were assigned to represent 1) a chosen position on the issue, 2) a statement given for or against Jack telling on Thomas, 3) a statement given in response to an anticipated objection, or a rebuttal, and 4) a repeat of the previously stated idea. Through this step, we eventually compiled a list of all propositions for and against Jack's telling on Thomas. The list was then consulted to assign a unique code to each distinct and acceptable reason advanced by students in their essays.

For example, consider the coding of the student essay presented in Table 3:

Table 3.

Coding a persuasive essay, ID 7834

Idea Unit	Free Node
[I think] he should not tell on Thomas.	Chosen Position: NO
[I think that because] maybe he might not have ever won anything before.	Reason 100
Also no one likes tattletales.	Reason 131
He helped a little.	Reason 108
Also, Jack feels sorry for him	Reason 129
and he is not very popular.	Reason 122
These are my reasons why Jack should not tell on Thomas.	Position-Repeat

We used different types of codes in order to distinguish between the statements supporting a positive vs. a negative position on the issue of whether Jack should tell on Thomas. Numbers from 0 to 99 were assigned to all statements supporting Jack telling on Thomas. Numbers from 100 up were used to label the statements opposing him telling. In the example above, the idea that nobody likes tattletales (underlined), which was frequently expressed by the students to support the position that Jack should *not* tell on Thomas, was coded using a free node *Reason 131*. A sum of all propositions numbered 100 and above for this student represents her ability to provide support for a chosen position. That is, this student has five supporting statements. She does not present any statements inconsistent with her chosen position (i.e., propositions coded with free nodes 0-99).

In the next step, we used QSR NVivo *Profile Coding for All Nodes* feature to produce initial summaries of student performance. This feature generates a report summarizing the number and types of nodes in all documents. The report can be easily exported into Excel, or other software program utilizing spreadsheets. The cells in Table 4 illustrate the results of this report, with the first row displaying the performance for the student from the previous example (ID 7834). Only few documents and nodes were selected for demonstration purposes.

Table 4

Summarizing The Nodes

ID	Chosen Position Yes	Chosen Position No	Reason 19	Reason 20	Reason 100	Reason 108	Reason 131	*Essay For*	*Essay Against*
7834	0	1	0	0	1	1	1	5	0
4751	1	0	1	0	0	0	0	7	2
1786	1	0	0	1	1	0	0	5	2

In the last step, we used Excel to create two summary measures of student performance, presented in the last two columns of Table 4. The first measure, *Essay-For*, represents the total number of propositions consistent with the chosen position. The second measure, *Essay-Against,* corresponds to the total number of opposing statements and rebuttals. Thus, we were able to separately measure student ability to support a chosen position (*Essay-For*) and to consider alternative perspectives (*Essay-Against*).

Two summary measures, *Essay-For* and *Essay-Against* were analyzed simultaneously using the Multivariate Analysis Of Variance (MANOVA) procedure, in order to see whether students in distinct treatment conditions differed on the

number of supporting and opposing statements generated in their essays. The omnibus MANOVA tests found statistically significant were followed up with univariate ANOVAs and multiple comparisons to further investigate the treatment differences. Results indicate that student performance on the persuasive essay was positively affected by participation in CR discussions. However, explicit instruction in argumentation did not produce expected gains, suggesting that performance can, at least initially, decline as explicitly-taught but as yet incompletely-learned principles undermine the ability to write an argument.

Relying on data-analytic procedures just described, Dong et al. (2006) evaluated persuasive compositions written by Chinese and Korean students. Researchers showed that participation in CR discussions resulted in the increased use of supporting and opposing reasons in written arguments, replicating our previous findings (Reznitskaya et al., 2007; Reznitskaya et al., 2001). Conceptual replications are crucial for supporting and extending theoretical principles. They are especially important for classroom research, as studies conducted under naturalistic conditions have numerous limitations due to the lack of experimental control. It is the sharing of useful research methodologies that creates opportunities for conducting replications using diverse settings and populations.

The second task in Reznitskaya et al. study (2007) was a recall of a persuasive text. The text was a 297-word passage about banning smoking in public places, with a clearly identifiable top-level structure. The text contained all parts of the argument that were explicitly taught to students, as well as organizational signals in connection with each part. For example, following the basic argument schema displayed in Figure 1, an objection to banning smoking in public was introduced with "Some people might say [OBJECTION]."

The first three steps in the analysis of text recalls were identical to the analysis of the essays. All text recalls were assigned a unique code, imported in QSR NVivo, and parsed into the idea units. In Step 4, we divided the original text into 33 distinct idea units, with each idea unit being assigned a unique code from 1 to 33. The list of idea units from the original passage was functionally comparable to the list of all acceptable and relevant statements used in scoring of persuasive compositions. Both lists served as templates against which student writing was evaluated.

Next, we compared the original text, divided into 33 idea units, to student recalls. If the recalled idea unit contained the same key terms and expressed the same meaning as an idea unit in the original text, it was assigned a Free Node corresponding to one of the 33 ideas from the original text. For example, unit 13 from the original text presented a health-related reason for banning smoking, stating that "Breathing in second hand smoke is risky". Below we present selected results from QSR Nvivo report of recalled text corresponding to the idea unit 13 from the original text:

Document '6277', 1 passages, 75 characters.

And breathing in other peoples cigarette smoke can make other people sick.

Document '748', 1 passages, 45 characters.

Breathing in second-hand smoke is also risky.

Document '9728', 1 passages, 52 characters.

Breathing in secondary smoke is very dangerous too.

Two summary measures, *Recall-For* and *Recall-Against* were generated in Excel. The *Recall-For* score represented students' ability to comprehend and recall statements supporting the main claim. The smoking-ban text presented two reasons why smoking should be banned in public. One reason was related to health problems; the other was concerned with environmental issues. Each reason was supported by several propositions that gave examples or cited scientific evidence. These propositions were termed elaborations. On the *Recall-For* measure, a student was given one point for recalling each of the following: the main claim; the first reason; the second reason; one or more elaborations of the first reason; and one or more elaborations of the second reason. The *Recall-Against* measure represented propositions advanced for the alternative perspective. The original text contained one elaborated counterargument and one elaborated rebuttal. Students received one point for recalling each proposition from the original text that expressed a counterargument, a rebuttal, or their elaborations.

Recall-For and *Recall-Against* variables were further analyzed in SPSS, using MANOVA. According to the analysis, recall of the argumentative text was generally insensitive to variations in treatments. Evidently, teaching students the schema in the context of oral discussion was not sufficient for the application of the schema in reading and recalling a persuasive text. We are currently conducting another study to further investigate the effects of oral argumentation on comprehension of written argument.

4. Towards Meaningful Quantification of Argumentation

In this paper, we demonstrated how rich verbal data can be interpreted with the use of theoretically-driven analytic schemes, software programs, and mathematical tools. The methodology described here has been successfully applied by researchers connected with CR group, who were able to examine, replicate, and extend important theoretical principles using diverse settings and populations. In this paper, we wanted to make our methodology available to a broader research community.

In Anderson et al. study (2001), we examined transition probabilities of argument stratagems and modeled the spread of stratagems to other children. This enabled us to systematically test the Snowball Hypothesis derived from social learning theories (e.g., Luria, 1981; Mead, 1962; Vygotsky, 1962) and further explicated through AST. Social learning theories continue to influence contemporary educators,

who call for the restructuring of traditional classroom practices (Billings & Fitzgerald, 2002; Keefer, Zeitz, & Resnick, 2000; Kuhn, 1992; Lipman, 1991). Yet, our understanding of the role that social interaction plays in argumentation development is quite limited. One reason for the scarcity of well-designed empirical studies of social influences is that many leading theories (e.g., Luria, 1981; Mead, 1962; Vygotsky, 1962) lack the desirable level of detail, explicitness, and clarity to guide research design and analysis (Anderson et al., 2001; Kucan & Beck, 1997; Webb & Palincsar, 1996; Wells, 1999; Wertsch & Bivens, 1992). Another reason is that there are few data-analytic approaches that are theoretically-driven, suitable for large datasets, and discussed in sufficient detail to allow for their use by future investigators.

In addition to substantive contributions, Anderson et al. study (2001) presented a detailed psychological theory of argumentation development that allows for generation of falsifiable predictions regarding the acquisition of argumentative discourse through group interactions. The theory guided the development and application of methodological procedures. The technology enhanced the sensitivity needed to capture important features of students' naturally occurring interactions and helped to accommodate large amounts of verbal data. It is through this fine-grained analysis of thousands of lines of children's discourse that we were able to identify and trace important trends in argumentation development.

In the second set of studies (Dong et al., 2006; Reznitskaya et al., 2007; Reznitskaya et al., 2001), we focused on assessing individual student performance, using such tasks as a persuasive composition and a recall of argumentative text. Many educators today strongly advocate the use of assessment instruments that allow for greater flexibility in responding (e.g., Baron, 1996; L. B. Resnick & Resnick, 1992; Strickland & Strickland, 1998). With the use of open-ended formats, "skills are displayed as a flow of performances and not as isolated behaviors out of context" (Millman & Green, 1988, p. 347). Proponents of open-ended tasks also argue that this format is more compatible with contemporary theories of learning, including cognitive, constructivist, and socio-cultural frameworks (e.g., Shepard, 2000). Open-ended questions should prompt students to engage in more complex and multifaceted behaviors, making this format better suited for measuring higher-level learning objectives (Grounlund, 1998; O'Neil, 1992; Wiggins, 1992).

Yet, many previous studies of argumentation development relied on fixed-choice tests of reasoning and related constructs, with items requiring students to *select* the answer from a list of alternatives (Daud & Husin, 2004; Fields, 1995; Wegerif, Mercer, & Dawes, 1999). Examples of commonly used fixed-choice measures include Cornell Critical Thinking Test (Ennis & Millman, 1985), Progressive Matrices Test (Raven & Court, 1963), and New Jersey Test of Reasoning Skills (Shipman, 1985). The use of fixed-choice format to measure argumentation and reasoning is especially problematic, as it obscures the thinking process that underlies the response (Chervin & Kyle, 1993; Halpern, 2003; Norris, 1991). "Multiple-choice tests…

provide only examinees' choices of answers to tasks, even though it is the reasoning that led to choices and not the choices themselves that are of greatest interest" (Norris, 1991, p. 459). Chervin and Kyle (1993) suggest that the "wrong" answer on a multiple-choice test of reasoning may be defensible, given specific audiences and frameworks for interpreting the question.

The continued reliance on fixed-choice tests of reasoning and related constructs may be due to the lack of theoretically-driven assessment methods. With open-ended formats, each student might react to the task requirements uniquely, thus creating challenges to consistency and efficiency of scoring. Although open-ended assessments are often described as less reliable and more resource-consuming than fixed-choice tests (e.g., Field & Brennan, 1989; Linn & Grounlund, 2000; Sax, 1997), their numerous potential advantages should prompt us to find the ways to address their limitations. Technological advances should help in resolving the issues of unreliability and inefficiency, even for larger datasets.

In our studies of argumentation (Dong et al., 2006; Reznitskaya et al., 2007; Reznitskaya et al., 2001), we employed open-ended tasks, such as writing a persuasive composition and reading a persuasive text. These activities preserve the authenticity of real-life situations involving the use of argumentation skills. They allow for shifting the emphasis from the ability to supply the correct answer to the process of arriving at a conclusion.

Software programs described in this paper helped us to enhance the quality and efficiency of the analysis. For example, QSR NVivo allowed for an on-going review of all instances of each assigned free node, across cases and raters, and within the context. This continuous review resulted in a more rigorous system and allowed to reduce subjectivity in judgment. The quality of the developed analytic framework was confirmed through high inter-rater reliability estimates for all summary measures, which ranged from $r=.87$ to $r=.92$.

We are currently working on further investigating the psychometric properties of our assessment tools (Reznitskaya, in preparation). Linn et al. (1991) caution against unquestionably accepting the purported benefits of performance-based assessments, arguing that merely changing to a less restricted response format does not, in itself, guarantee that the inferences will be more valid than those obtained with traditional fixed-choice items. Thus, evidence must be collected to support interpretations of student performance (Linn, et al. 1991). For example, our evaluation strategy for persuasive compositions involved counting the number of propositions supporting and opposing a taken position. Alternative scoring frameworks can reflect different dimensions of performance and deepen our understanding of argumentation as a construct. For instance, the quality of the proposed reasons can be taken into account through assigning differential weightings to student propositions. In their study of children's arguments, Means and Voss (1996) proposed a hierarchy of reasons, suggesting, for example, that appealing to direct consequences of a given action is better than appealing to authority or to personal experience.

Also, the coding system described here did not take into consideration the relationships between the propositions comprising an argument. In other words, the manner in which the statements were linked was not assessed. Analyzing argument cohesion is another interesting direction to take in future studies.

In this paper, we discussed theoretical and data-analytic frameworks used to examine important features of argumentative discourse. Gathering rich verbal data and then transforming it into numerical form can bring about many benefits typically associated with numbers, as opposed to words. These benefits include increased precision and differentiation, reduced disagreement, efficiency of communication, and the ability to use mathematical modeling for discovering useful generalizations. However, these advocated advantages of quantification are lost when we abandon such scientific ideals as search for meaning, appreciation of complexity, and attention to detail. Data-analytic strategies described in this paper allowed us to maintain the tension between the quantitative vs. qualitative extremes of scientific inquiry and to reconcile the artificial divide between them.

References

Almasi, J. F., O'Flahavan, J. F., & Arya, P. (2001). A comparative analysis of student and teacher development in more or less proficient discussions of literature. *Reading Research Quarterly, 36*(2), 96-120.

Alveraman, D. E., O'Brien, D. G., & Dillon, D. R. (1990). What teachers do when they say they're having discussions of content area reading assignments: A qualitative analysis. *Reading Research Quarterly, 25*, 297-322.

Anderson, R. C. (1977). The notion of schemata and the educational enterprise. In R. C. Anderson, R. J. Spiro & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 415-431). Hillsdale, NJ: Erlbaum.

Anderson, R. C., Chinn, C., Waggoner, M., & Nguyen, K. (1998). Intellectually stimulating story discussions. In J. Osborn & F. Lehr (Eds.), *Literacy for all: Issues in teaching and learning* (pp. 170-186). New York: Guilford.

Anderson, R. C., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., et al. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and instruction, 19*(1), 1-46.

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior, 17*, 1-12.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685-730.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays by M. M. Bakhtin*. Austin, TX: University of Texas Press.

Bakhtin, M. M. (1986). *Speech genres and other late essays* (V. W. McGee, Trans.). Austin, TX: University of Texas Press.

Baron, J. B. (1996). *Performance-based student assessment: challenges and possibilities*. Chicago: University of Chicago Press.

Billings, L., & Fitzgerald, J. (2002). Dialogic discussion and the Paideia Seminar. *American Educational Research Journal, 39*(4), 907-941.

Bohstedt, J. (1994). The dynamics of riots: Escalation and diffusion/contagion. In M. Portegal & J. F. Knutson (Eds.), *The dynamics of aggression: Biological and social processes in dyads and groups* (pp. 257-306). Hillsdale, NJ: LEA.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*, 207-230.

Chambliss, M. J. (1995). Text cues and strategies successful readers use to construct the gist of lengthy written arguments. *Reading Research Quarterly, 30*(4), 778-807.

Chambliss, M. J., & Murphy, P. K. (2002). Fourth and fifth graders representing the argument structure in written texts. *Discourse Processes, 34*(1), 91-115.

Cheng, P., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391-416.

Chervin, M. I., & Kyle, J. A. (1993). Collaborative inquiry research into children's philosophical reasoning. *Analytic Teaching, 13*(2), 11-32.

Chinn, C. A., & Anderson, R. C. (1998). The structure of discussions that promote reasoning. *Teachers College Record, 100*(2), 315-368.

Daud, N. M., & Husin, Z. (2004). Developing critical thinking skills in computer-aided extended reading classes. *British Journal of Educational Technology, 35*(4), 477-487.

Dong, T., Anderson, R. C., Li, Y., & Kim, I. (2006). *Language of home and school: Discourse mismatch reconsidered*. Champaign: University of Illinois, Center for the Study of Reading.

Ennis, R. H., & Millman, J. (1985). *Cornell Critical Thinking Test*: Critical Thinking Books & Software.

Field, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational Measurement*. New York: Anerican Council on Education.

Fields, J. I. (1995). Empirical data research into claims for using philosophy techniques with young children. *Early Childhood Development and Care, 107*, 115-128.

Freedman, A., & Pringle, I. (1988). Why students can't write arguments. In N. Mercer (Ed.), *Language and literacy from an educational perspective* (Vol. 2: In Schools, pp. 233 - 242). Milton Keynes: Open University Press.

Govier, T. (1985). *A practical study of argument*. Belmont, CA: Wadsworth Publishing Company.

Grounlund, N. E. (1998). *Assessment of student achievement*. Boston, MA: Allyn & Bacon.

Halpern, D. F. (2003). The "how" and "why" of critical thinking assessment. In D. Fasko (Ed.), *Critical thinking and reasoning* (pp. 331-354). Cresskill, NJ: Hampton Press, Inc.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88-140.

Holmes, E. T. (1977). *Amy's goose*. New York: Crowell.

Keefer, M. W., Zeitz, C. M., & Resnick, L. B. (2000). Judging the quality of peer-led student dialogues. *Cognition and Instruction, 18*(1), 53-81.

Kim, I., Anderson, R. C., Nguyen-Jahiel, K., & Archodidou, A. (in submission). Discourse patterns during children's collaborative online discussions. *Journal of the Learning Sciences*.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco, CA: Jossey-Bass.

Kucan, L., & Beck, I. L. (1997). Thinking Aloud and Reading Comprehension Research: Inquiry, Instruction, and Social Interaction. *Review of Educational Research, 67*(3), 271-299.

Kuhn, D. (1991). *The skill of argument*. Cambridge, UK: Cambridge University Press.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*(2), 155-177.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and instruction, 15*(3), 287-315.

Kumar, S., & Miller, K. F. (in press). Let SMIL be your umbrella: Software tools for transcribing, coding, and presenting digital video in behavioral research. *Behavior Research Methods, Instruments, & Computers*.

Lefebvre, L., Whittle, P., Lascaris, E., & Finkelstein, A. (1997). Feeding innovations and forebrain size in birds. *Animal Behaviour*(53), 549-560.

Li, Y., Anderson, R. C., Nguyen-Jahiel, K., Dong, T., Archodidou, A., Kim, I., et al. (2007). Emergent leadership in children's discussion groups. *Cognition and Instruction, 25*(1), 75-111.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

Linn, R. L., & Grounlund, N. E. (2000). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Merrill.

Lipman, M. (1991). *Thinking in education*. Cambridge: Cambridge University Press.

Luria, A. R. (1981). *Language and cognition*. New York: Wiley.

Mayer, R. E. (1985). Structural analysis of science prose: Can we increase problem solving performance? In B. K. Britton & J. B. Black (Eds.), *Understanding of expository text* (pp. 65-87). Hillsdale, NJ: Erlbaum.

McCann, T. M. (1989). Student argumentative writing knowledge and ability at three grade levels. *Research in the Teaching of English, 23*(1), 63-77.

Mead, G. H. (1962). *Mind, self, and society from the standpoint of a social behaviorist*. Chicago: University of Chicago press.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14*(2), 139-178.

Meyer, B. J., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 1*, 72-103.

Millman, J., & Green, J. (1988). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational measurement* (third ed.). New York: American Council of Education.

Mishra, P., & Brewer, W. F. (2003). Theories as a form of mental representation and their role in the recall of text information. *Contemporary Educational Psychology, 28*, 277-303.

Morris, M. (Ed.). (1994). *Epidemiology and social networks: Modeling structured diffusion*. Thousand Oaks, CA: Sage Publications.

NAEP. (1994). *The National Assessment of Educational Progress 1992 Report Card*. Princeton, NJ: Educational Testing Service.

NAEP. (1999). *National Assessment of Educational Progress Writing Report Card for the Nation and the States*. Retrieved December 8, 2004, from http://nces.ed.gov/nationsreportcard//pdf/main1998/1999462.pdf

NAEP. (2002). *The nation's report card: Reading 2002*. Retrieved December 8, 2004, from http://nces.ed.gov/nationsreportcard/pdf/main2002/2003521.pdf

Norris, S. (1991). Informal reasoning assessment: Using verbal reports of thinking to improve multiple-choice validity. In F. J. Voss, D. N. Perkins & J. W. Segal (Eds.), *Informal reasoning and education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nystrand, M., Wu, L., Garmon, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes, 35*(2), 135-200.

O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership, 49*(8), 14-19.

Pellegrini, A. D., Galda, L., & Rubin, D. (1984). Persuasion as a social-cognitive activity: The effects of age and channel of communication on children's production of persuasive messages. *Language and Communication, 4*(4), 285-293.

Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart, & Winston.

QSR. (1997). QSR NUD*IST 4 [Computer software]. Victoria, Australia: Qualitative Solutions and Research.

QSR. (1999). QSR Nvivo [Computer software]. Victoria, Australia: Qualitative Solutions and Research.

Raven, J. C., & Court, J. H. (1963). *Progressive matrices*. Oxford, UK: Oxford Psychologists Press.

Resnick, L., Salmon, M., Seitz, C. N., Wathen, S. H., & Holowchack, M. (1993). Reasoning in conversation. *Cognition and Instruction, 11*, 347-364.

Resnick, L. B., & Resnick, D. P. (1992). Assessing thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction. Evaluation in education and human services* (pp. 37-75).

Reznitskaya, A., & Anderson, R. C. (2002). The argument schema and learning to reason. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction* (pp. 319-334). New York: Guilford.

Reznitskaya, A., Anderson, R. C., & Kuo, L. (2007). Teaching and learning argumentation. *Elementary school journal, 107*(5).

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*(2 & 3), 155-175.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

Roth, W. M. (1996). Knowledge diffusion* in a Grade 4-5 classroom during a unit on civil engineering: An analysis of a classroom community in terms of its changing resources and practices. *Cognition and Instruction, 14*(179-220).

Sax, G. (1997). *Principles of educational and psychological measurement*. Belmont, CA: Wadsworth Publishing Company.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 1-14.

Shipman, V. C. (1985). *New Jersey Test of Reasoning Skills*. Montclair, NJ: Institute for the Advancement of Philosophy for Children.

Strickland, K., & Strickland, J. (1998). *Reflections on assessment*. Portsmouth, NH: Boynton/Cook Publsihers.

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

Toulmin, S. E., Rieke, R., & Janik, A. (1979). *An introduction to reasoning*. New York: Macmillan.

van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Orlando, FL: Academic Press.

van Eemeren, F. H., & Grootendorst, R. (1984). *Speech acts in argumentative discussions*. Foris: Dordrecht and Cinnaminson.

van Eemeren, F. H., Grootendorst, R., & Henkenmans, F. S. (1996). *Argumentation: Analysis, evaluation, presentation*. Hilldale, NJ: Erlbaum.

Voss, F. J., & Means, M. L. (1991). Learning to reason via instruction in argumentation. *Learning and Instruction, 1*, 337-350.

Vygotsky, L. S. (1962). *Thought and Language*. Cambridge: MIT Press.

Vygotsky, L. S. (1981). The genesis of higher-order mental functions. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 144-188). Armonk, NY: Sharpe.

Waggoner, M., Chinn, C. A., Yi, H., & Anderson, R. C. (1995). Collaborative reasoning about stories. *Language Arts, 72*, 582-589.

Walton, D. (1996). *Argument structure: A pragmatic theory*. Toronto: University of Toronto Press.

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841-873). New York: Simon & Schuster Macmillan.

Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: An empirical investigation of a possible sociocultural model of cognitive development. *Learning and Instruction, 9*(6), 493-516.

Wells, G. (1999). *Dialogic inquiry: Toward a sociocultural practice and theory of education*. Cambridge, UK: Cambridge University Press.

Wertsch, J. V. (1985). *Vygotsky and social formation of mind*. Cambridge, MA: Harvard University Press.

Wertsch, J. V., & Bivens, J. A. (1992). The social origins of individual mental functioning: Alternatives and perspectives. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 14*(2), 35-44.

Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership, 49*(8), 26-34.

Yeh, S. (1998). Empowering education: Teaching argumentative writing to cultural minority middle-school students. *Research in the Teaching of English, 33*, 49-81.

Correspondence concerning this article should be addressed to:

Alina Reznitskaya
Montclair State University
University Hall 2193
Montclair, NJ 07043

e-mail: reznitskayaa@mail.montclair.edu
phone: (973) 655-4080
fax: (973) 655-6915