Instructional influences on English language learners' storytelling

Shufeng Ma[1], Richard C. Anderson[1], Tzu-Jung Lin[2],

Jie Zhang[3], Joshua Morris[1], Kim Thi Nguyen-Jahiel[1], Brian Miller[4], May Jadallah[5],

Theresa Scott[6], Jingjing Sun[7], Kay Grabow[8], Beata Latawiec[9], Fang-Hsien Yi[1]

[1]*University of Illinois at Urbana-Champaign*

[2]*Ohio State University*

[3]*Western Kentucky University*

[4]*Towson University*

[5]Illinois State University

[6]*Washington Elementary School, Champaign, Illinois*

[7]University of Montana

[8]*Thomas Paine Elementary School, Urbana, Illinois*

[9]*Wichita State University*

Author note:

The research reported in this paper was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080347 to the University of Illinois at Urbana-Champaign, Richard C. Anderson, Principal Investigator. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. The authors are pleased to acknowledge the contributions of Nikisha Blackmon, Aini Marina Ma'rof, and Xiaoying Wu.

Corresponding authors:

Shufeng Ma and Richard C. Anderson
Center for the Study of Reading
University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, IL 61820
217-333-5845 (voice) 217-244-4501 (fax)
sma10@illinois.edu
csrrca@illinois.edu

ABSTRACT

Instructional influences on storytelling were investigated among 210 Spanish-speaking fifth graders. Participants received a 6-week socio-scientific unit involving collaborative group work or direct instruction, or were in control classes that continued regular instruction. Then students individually told a story prompted by a wordless picture book. Analysis of story transcripts indicated that the stories produced by students who had participated in collaborative groups were more fully elaborated and cohesive. Students who had interacted in collaborative groups generated significantly longer chains of reasoning (many 5-7 link chains) than students who had received direct instruction (mostly 1-2 link chains). The results suggest collaborative group work may be an effective instructional approach to foster ELL's oral narrative skill and causal reasoning.

***Keywords*:** English language learner, Collaborative group work, oral narrative, communicative competence, Multi-link causal reasoning

In the 2008-2009 academic year, the number of registered English Language Learners (ELLs) from pre-K to grade 12 in United States was 5.35 million (Padolsky, 2011). Nearly 80% of ELLs are children of Spanish-speaking immigrants from Central and South American countries (Kindler, 2002). In addition to limited English proficiency, large percentages of Latina/o immigrants have lower socioeconomic status and educational levels than other minority groups (Goldenberg, 1996; Candelaria & Llorente, 2009). ELLs have lower academic achievement in reading and mathematics compared to non-ELLs. According to the 2012 National Assessment of Educational Progress, fourth-grade Hispanic students scored 25 points lower than non-Hispanic white students in reading and eighth-grade Hispanic students scored 22 points lower. Influential factors that impede ELL's literacy growth include low social economic status, limited home literacy resources, poor oral English proficiency, and lack of vocabulary knowledge (Snow, Burns, & Griffin, 1998). Among these factors, oral English proficiency has the most profound relationship with ELL's academic attainment (Genesee, Lindholm-Leary, Saunders, Christian, 2005). It has been estimated that half of ELLs with low English proficiency drop out of high school. However, for ELLs who can speak English well, the high school completion rate is 82% (August & Shanahan, 2008).

It is apparent that there is an urgent need for effective intervention programs to improve ELL's oral English proficiency. This paper reports a study of the effects of two contrasting instructional approaches on the oral English development of ELLs, direct instruction and collaborative group work. The following sections [A] review previous research on the impact of different instructional approaches on language development, [B] describe methods of assessing oral language proficiency, and [C] provide the rationale for the present study.

Instructional methods and language development

Previous research has explored instructional approaches to promote ELL's language proficiency. Teacher-guided, direct instruction has been found to be effective when implemented well. With teachers' explicit instruction, students can develop fundamental language and literacy skills such as phonological awareness, decoding, and basic vocabulary (Avila & Sadoski, 1996; Kamps et al., 2007, Kelcey & Carlisle, 2013). Direct instruction can also promote strategic learning; for example, teachers can impart strategies for vocabulary learning and text comprehension (Genesee et al., 2005; Pressley & el Dinary, 1997). However, the effectiveness of direct instruction depends on curriculum design (Stein, Carnine, & Dixon, 1998). A loosely-organized curriculum over-packed with concepts to be covered may lead to rote learning rather than knowledge construction. According to Stein and her colleagues (1998), well-designed direct instruction emphasizes integration of skills and concepts, scaffolded instruction, explicitly taught strategies, a balance of highlights and details, and systematic review.

Besides direct instruction, a family of interactive methods, with names such as Thinking Together, Shared Inquiry, Exploratory Talk and Instructional Conversations, has proven to enhance classroom interaction and produce gains on several types of outcome measures (Murphy et al., 2009). For example, Exploratory Talk, in which students make explicit reasoning and constructively develop each other's ideas, has shown positive results in improving British and Mexican children's individual reasoning and problem solving (Rojas-Drummond & Mercer, 2003). Instructional Conversations, proposed by Tharp and Gallimore (1989) and Goldenberg (1992) to improve teacher-student interaction, was found to promote ELLs' higher-order thinking (Saunders & Goldenberg, 1996).

The present study features Collaborative Reasoning (Anderson, Chinn, Waggoner, & Nguyen-Jahiel, 1998), an interactive alternative to direct instruction that has shown promise in changing the quality of classroom talk (Chinn, Anderson, & Waggoner, 2001) and improving educational outcomes (Murphy et al., 2009; Reznitskaya et al., 2009). Collaborative Reasoning (CR) discussions provide an open forum for students and minimize the dominant role of the teacher, which not only creates an interactive learning environment, but also facilitates communication monitoring as students learn to make contextually appropriate contributions to discussions and to ask for clarification when they don't understand. Small heterogeneous discussion groups are composed, balancing students' talkativeness, reading level, and social status among peers. Students read a story individually, then come together to discuss a ‘big question’ about a dilemma in the story. Students are encouraged to take their own positions on the big question. The participants take turns to talk without raising hands and manage the group discussion themselves (Anderson et al., 1998). They are encouraged to freely participate and challenge others’ opinions. They have a chance to express extended ideas in public and organize reasoning and evidence to support their positions. In comparison to typical forms of classroom discussion, students’ rate of talk almost doubles during CR and the talk more frequently involves elaborating text propositions, making predictions, using evidence, asking for and providing clarification, and expressing and considering alternative perspectives (Chinn, Anderson & Waggoner, 2001).

Reznitskaya and colleagues (2009) summarized research about group dialogue and individual outcomes during Collaborative Reasoning discussions. They concluded that CR discussions helped students develop the skills of argument to address complicated problems. In studies in China and Korea, as well as the United States, students who participated in CR wrote essays that contain more acceptable arguments, counterarguments, and rebuttals than control students, which Reznitskaya et al. explain in terms of ‘argument schema theory.’ CR discussions enhanced students’ abstract knowledge of argumentation, which consequently enabled students to learn “how to think as well as what to think” (Reznitskaya et al., 2009).

Based on the promising but preliminary results with the Collaborative Reasoning, the present study aimed to investigate the effects of CR, in combination with other collaborative activities, on ELLs’ basic oral language proficiency, communicative competence, and high-order cognitive skills.

Assessment of oral language proficiency

Assessment of English language learners often focuses on the basic elements of language proficiency (August & Shanahan, 2008), but neglects *communicative competence* (Hymes, 1972; Cazden, 2011), so many language educators advocate testing ability to communicate in life-like situations (August &

Shanahan, 2008; Baker, 2011). Oral narrative ability is at the core of communicative competence and an influential factor predicting reading ability and school achievement (Genesee et al., 2005; Miller et al., 2006).

Various features of language production can be coded from oral narratives (Miller & Chapman, 2010) yielding, for example, measures of syntactic complexity and the frequency of ‘mazes’ (e.g. false starts, pauses filled with ‘um’). Narratives may be elicited from children using a wordless picture book, following a standardized protocol for introducing the task and prompting children when needed, for instance, if they stop short of providing a complete story (Berman & Slobin, 1994). Oral narrative assessment includes the evaluation of story quality as well as features of language production. The narrative scoring framework proposed by Pearson (2002) is typically used to rate story quality. The framework is based on Stein and Glenn’s (1979) story schema theory, which encompasses the elements of a well-formed story.

According to Stein and Glenn (1979), a story is composed of two essential components—the setting and the episode system. The function of the setting is to introduce the main characters in the story and describe the “social, physical, or temporal context in which the remainder of the story occurs” (Stein & Glenn, 1979, p. 60). The episode system represents the rest of the narrative, which is usually a collection of different episodes. Each episode is composed of internal responses and external responses. Internal responses refer to a character’s feelings, thoughts, and goals. External responses are the character’s actions in the circumstances. The consequences of internal and external responses motivate main character’s reactions to the event—either continue to make attempts or reach an end state when the goal is achieved. The organization of fundamental elements of a story is the key indicator of children’s communicative competence.

Trabasso and van den Broek (1985) theorized that events in a narrative are organized as a causal network. All the elements in story schema theory, except for *setting* and *end state*, serve as antecedent and consequent events in a causal chain. The importance of an event is determined by its relationship with other events and its position in the hierarchical structure. According to Magliano (1999), the causal network model explains how story elements are bound together based on the different types of causal relationships between episodic categories. *Goals* can activate *attempts* and *attempts* can result in *outcomes*.

Causal connections are constructed by storytellers to specify the relationships between events. A given event relies on the occurrence of previous events and sets up subsequent events. We have termed the ability to organize information and bridging inferences into coherent causal chains *multi-link causal reasoning* (Lin, Ma, et al., 2011). In storytelling, multi-link causal reasoning is reflected by the ability to link a sequence of events together. Children's ability to generate multi-link reasoning chains is essential for story understanding and production and an indication of higher-order cognitive skill.

Rationale for this study

This study explored the effects of contrasting instructional approaches on ELLs' language production, communicative competence, and higher-order thinking and reasoning as revealed in their oral narratives. One instructional approach was Collaborative Group work (CG), which combined Collaborative Reasoning and other group activities. This approach was compared to teacher-led whole-class Direct Instruction (DI) and regular instruction in control classrooms.

Positive effects from direct instruction are evident in the early stages of schooling when receptive learning can contribute to English language foundations (Genesee et al, 2005). However, at later stages, direct instruction may restrict opportunities for language use in social and communicative circumstances. Classroom talk during direct instruction ordinarily takes place in the question-response-evaluation format. The interaction starts with a question asked by the teacher, proceeds with students' response to that question, and ends with an evaluation from the teacher. The question-response-evaluation format makes it difficult for ELLs to produce extended talk (Arrega-Mayer & Perdomo-Rivera, 1996). Students' thinking is constricted because students have to follow the teacher's logic rather than initiate their own thinking.

Students have limited control over when they can speak, little or no say about the topic of discussion, and negligible authority to evaluate whether contributions are acceptable (Wells & Arauz, 2006). In an observational study of 145 third- to fifth-grade classrooms in 20 low-income schools with large enrollments of ELLs, McCaslin and her associates (2006) reported that direct, teacher-led instruction predominated in virtually every classroom. Nearly 75% of instructional time in these classrooms focused on fundamental facts, basic skills, content learning, along with modest levels of elaboration and related thinking. Only 3% of instructional opportunities were devoted to higher-order thinking and reasoning. The few questions asked by students were mainly concerned with task procedures and correctness of answers; only 3% of student questions were reported to involve thinking or knowledge exploration (McCaslin et al, 2006).

Because of the constraints on spontaneous language and extended thinking associated with direct instruction, interactive learning approaches are often proposed as an alternative or supplement to facilitate development of oral language proficiency (cf. Genesee et al., 2005; Ellis, 2005). Theoretically, children's language and thinking develop in dialogic interaction during collaborative group work through several socialization mechanisms, including observing and emulating other children (Bandura, 1986; Lin et al., 2012), assistance less-competent children receive from others (Vygotsky, 1978; Webb & Mastergeorge, 2000), and the stimulation all of the children experience while resolving sociocognitive conflicts (Piaget, 1976/1947; Johnson & Johnson, 2009).

Several studies suggest that collaborative, interactive approaches can improve ELLs' language proficiency (e.g. Rojas-Drummond & Mercer, 2003; Saunders & Goldenberg, 2007), although there have also been disappointments (see Gersten & Baker, 2000, p. 461). It is important to emphasize that interaction that enables children to develop the conversational skills for informal social situations is unlikely to be sufficient to impact the children's school performance (Cummins, 1986). Language for social situations is easier in several respects than 'academic language.' Snow (2014) explains that "features of academic language include sophisticated vocabulary forms, explicit discourse markers (e.g. *nonetheless,* *therefore*), information packing through the use of nominalizations, embedded relative clauses, and subjectless passives. . . [These features] constitute an enormous challenge to struggling readers, second-language readers, and to those who have not been inducted into the use of academic language in oral contexts (p. 120)."

Zhang, Anderson, and Nguyen-Jahiel (2013) studied whether the Collaborative Reasoning approach could help ELLs improve English reading, listening, speaking, and writing. They found that CR discussions of ethical and practical dilemmas raised in stories accelerated Spanish-speaking ELLs' oral and written English, as well as their motivation, engagement in discussions, and English learning attitudes. A weakness of this study was the small sample so as the authors say the findings were "suggestive rather than definitive."

This study extended Zhang and colleagues (2013) with a larger sample, improved methods, and in a context that stressed academic language. The study compared the storytelling of Hispanic American fifth graders who completed a six-week unit on wolf reintroduction and management using Collaborative Groups or Direct Instruction, or who continued regular instruction. Following the unit on wolves, among other tasks, the children told a story prompted by a wordless picture book to provide an authentic evaluation of their oral English proficiency. The stories were coded for several measures of basic language production, communicative competence, and thinking and reasoning. Based on the previous findings that Collaborative Reasoning approach substantially increased the quantity and quality of classroom talk as compared to the direct instruction (Chinn, Anderson & Waggoner, 2001), the CG approach was expected to lead to gains in all three aspects, because in collaborative groups children have more opportunities for high quality interaction. It was anticipated that the DI approach might also lead to gains, as compared to the control condition, because as implemented in this study DI involved richer concepts and greater use of connected academic language than is typical in classrooms containing large numbers of ELLs.

Method

Participants

Included in the analysis were 210 Hispanic American fifth graders from 18 classrooms in 4 elementary schools in a city in northern Illinois. Spanish was the first language of all of the participants. English was the primary language of school instruction. More than 80% of the children were registered for free or reduced price lunch. The sample was balanced in gender (Girl: N=103; Boy: N=107). The average age of the participants was 10.8 (SD = 0.4).

Before the intervention, a home literacy survey was completed by parents. The survey asked for information about the language children spoke at home, parent years of education, the language of instruction in the first grade, and home literacy resources and practices. The results showed that most parents (father: 87%, mother: 83%) had a high school level of education or lower. Most children (84%) communicated with their parents in Spanish or an equal combination of Spanish and English. Language use in the first grade is an indicator of early or late English acquisition; two thirds of the children had instruction in Spanish or a combination of Spanish and English in the first year of school. Most families had limited literacy resources; 78% of families reported fewer than 40 child books and 75% of families had fewer than 40 adult books. Nearly half of the parents reported telling stories to children once a month or never. Most parents (61%) said they did not read books with children frequently and 14% of parents said they had never read books with their children. Twenty percent of the parents indicated that they never went to the library with their child. Half of the parents reported they did not spend much time themselves (at least once a week) reading books, newspapers, or magazines at home.

Design and procedure

Triples of classes matched on demographic characteristics and average previous-year performance on the Illinois Standards Achievement Test (ISAT) were randomly assigned to three intervention conditions for a curriculum unit on wolf reintroduction and management. The intervention conditions were collaborative group work (CG), whole-class Direct Instruction (DI), or wait listed control condition. Classroom ISAT performance was initially reported to us by school principals and later confirmed when individual student scores became available. Table 1 presents demographic characteristics, indicators of Spanish use, and average scores on tests of reading and language in the 18 classrooms in this study. There were no significant differences among the three conditions on any of these measures.

[Insert Table 1 here]

The intervention was conducted in two waves across two academic years and each wave had 18 fifth-grade classrooms. Altogether the project included over 800 students recruited from schools serving low-income neighborhoods and predominantly students from two underserved populations—African Americans and Hispanic Americans. This study focused on the Hispanic American students from the 18 classrooms with predominant Hispanic American enrollment, nine classrooms from each wave.

In the first wave, the Hispanic American students in each intervention condition were enrolled in one bilingual classroom with English-Spanish instruction and two mainstream classrooms with instruction entirely in English. Students were assigned to classrooms based on their performance on the annual statewide English proficiency test in Illinois. Those below the cut-score were assigned to sheltered bilingual classrooms; those above the cut-score were instructed in mainstream classrooms. Before the second wave, the participating schools abandoned the bilingual program for middle grade students, so all of the students were in mainstream classrooms, although it was reported that assistance in Spanish was still occasionally provided. Participant observers rarely saw any use of Spanish in the second wave, and not much use in the first wave even though the students least proficient in English were in sheltered bilingual classes. The infrequent use of Spanish reflected school district policy which emphasized acquiring proficiency in English rather than proficiency in both languages.

Pooling over the two waves, and setting aside students with other ethnic backgrounds, there were 324 Spanish-speaking ELLs in the study, among which 210 students received the oral narrative assessment. These students were approximately 65% of the Hispanic American students in the participating classrooms. The remainder could not be given the individual oral assessment because of limitations on time and resources. To determine which students would receive the assessment, each assistant was given a list of students with randomly ordered 'target students' on the top and the remainder of the class also randomly ordered below. Target students were so called because the video camera was trained on them throughout the Wolf Unit. Target students were selected with the help of the teacher to be a representative cross-section of the class in terms of academic level, talkativeness, gender, and ethnicity. To avoid interrupting classes all day long and moving back and forth between different schools, the assessment was conducted in one class at a time. Target students in each classroom received the assessment first, then according to the randomized list as many additional children as could be accommodated in the available time were assessed. The analyses reported in this paper involved 70 Hispanic American students in the CG condition, 68 in the DI condition, and 72 in the control condition.

Comparing students who received the storytelling assessment and those who did not, the demographic characteristics, percentage of children who speak Spanish with parents, and percentage who used English in the first grade were all very similar. Separate ANOVAs were performed with Gates-MacGinitie reading score, rapid naming speed, and fourth-grade ISAT reading score as dependent variables, intervention condition and whether students received the oral assessment or not as fixed effects, and classroom as a random variable. The results showed that students who received the storytelling assessment had significantly higher Gates-MacGinitie reading scores than those who did not, F (1, 303) = 4.40, p = .037; however, there was no significant condition difference, F (2, 303) = 0.65, p = .52, or interaction between condition and whether students received the oral assessment or not, F (2, 303) = 0.02, p = .98. This result indicates that selection favored students who had higher reading comprehension scores, but this applied to all conditions. Students who received the oral assessment and students who did not receive the oral assessment showed equivalent performance on the rapid automatized naming task, F

(1, 303) = 1.11, p = .29. Among the students who took the ISAT reading test in fourth grade, no difference was found between students who received the oral assessment and those who did not, F (1, 219) = 1.66, p = .20. Therefore, although selection of students to receive the storytelling task was associated with reading comprehension, the students were alike in other respects and characteristics of selected students were not differentially related to instructional condition.

Wolf Reintroduction and Management Unit

The study involved a six-week long Wolf Reintroduction and Management Unit, which is constructed around a science and public policy issue in an imaginary community. The people of Winona County are concerned about a pack of wolves that has been sighted nearby. Fears for the safety of children and pets, as well as potential threats to ranching and tourism, lead the County Board to write a letter to the Wolf Management Agency asking permission to hire professional hunters to kill the wolves. Students role played being the officials in the agency who must make the decision about eradicating the wolves.

As Jadallah et al. (2009) explained, the Wolf Unit uses a variety of information sources to help students learn about issues surrounding wolf reintroduction and management. Students read texts that incorporate different genres (e.g., expository text, newspaper articles, and formal letters). The unit integrates language arts, math, science, and social science. The Wolf Unit covers three domains of knowledge—ecosystem, economy and public policy. Each knowledge domain represents one thread of thinking on the relationship between wolves and the world around them. Reading materials and activities in each domain have an argumentative structure and cover both sides of the issues.

Intervention conditions

Collaborative group work (CG) was a combination of Collaborative Reasoning discussions and other group activities. After an introduction to Winona and its problem with wolves, students were broken into groups to discuss the 'big question'—whether Winona should be permitted to hire professional hunters to the kill wolves. On a typical day during the unit, the task for a small group was to answer a sub-question related to the big question, for example, “What effect would killing the wolves have on the elks?” Groups worked independently and spoke freely among themselves, with occasional assistance from the teacher. Each small group was assigned to become ‘experts’ in one of the three domains of knowledge (ecosystem, economy, public policy). After four weeks of group work, the children in each expert group shared what they had learned in a poster presentation to the whole class. Then new discussion groups were created, with members from the three different expert groups, to reconsider the big question in a Collaborative Reasoning discussion. As the last activity in the unit, students independently wrote a policy decision letter on whether killing the wolf pack should be permitted.

Direct Instruction (DI) entailed teacher-guided whole-class activities and individual seatwork. Students in DI condition sat facing toward the teacher. Students were supposed to raise their hands and wait for the teacher to select them before speaking. The teacher led students through all three domains of knowledge in the Wolf Unit. Activities that were completed in small groups in CG classrooms were completed individually as seatwork in DI classrooms. DI students discussed the policy decision as a whole class. Finally, as in the CG condition, students independently wrote a decision letter on whether killing the wolves should be allowed.

Wait-listed control classes continued to receive regular language art instruction during the intervention period. Control classes had the opportunity to study the Wolf Unit in the following semester.

Teachers who implemented CG or DI interventions attended a two-day workshop to receive a detailed introduction to the Wolf Management Unit, discuss the design and content of the curriculum, and receive training in the method to which they were assigned. Teachers watched videos of Wolf Unit as it had been implemented in other classrooms taught by teachers using the methods they were supposed to use. Teachers who implemented collaborative group work learned about the goal of the intervention, the research and theory supporting Collaborative Reasoning, how to facilitate CR discussions and effective strategies for promoting group work. Teachers who implemented whole-class direct instruction learned about the research and theory supporting explicit teaching of concepts and strategies, and effective strategies for direct instruction. The workshop staff included elementary school teachers known for their expertise in collaborative group work or direct instruction.

A research assistant was assigned to every classroom as a participant observer to administer tests, video record lessons, take field notes, and assist the teacher. The field notes written by the research assistants left no doubt that the Wolf Unit was implemented in every classroom and did not suggest significant departures from the assigned instructional approach, although from time to time partner activities among the students were observed in some DI classrooms and most CG teacher occasionally explained concepts to the whole class.

To further gauge implementation of the Wolf Unit and fidelity to the assigned instructional approach, 146 four-minute episodes were systematically sampled from the roughly 500 hours of videos of Wolf Unit lessons in the CG and DI classrooms recorded during the present study [lessons in Control classrooms were not video recorded]. Transcriptions of the episodes were searched for uses by teachers and students of any of 76 academic vocabulary items listed in the Wolf Unit glossary. The items consisted of 26 domain-specific vocabulary words [e.g. *omnivore, food chain]* and 50 general academic vocabulary words [*economy, balance*], as classified by the Common Core State Standards (Common Core State Standards Initiative, 2010, p. 33). If participants were on-task and actually studying the Wolf Unit, uses of academic vocabulary should be evident. If participants played the roles assigned to them, teacher use of the terms should be higher in DI classrooms and student use of the terms higher in CG classrooms.

Assessments

Before the intervention, the Gates-MacGinitie (MacGinitie, MacGinitie, Maria, & Dreyer, 2000) reading comprehension test was administered. The test entails answering multiple-choice questions after reading short passages. The raw score was corrected for guessing, which improved reliability and predictive validity. Also before the intervention, students individually completed a rapid automatized naming task (Snodgrass & Vanderwart, 1980) to assess basic oral English proficiency. Students named common objects, such as bike, rabbit, and bus, in two sets of pictures. Students were asked to name the objects as quickly as they could but were allowed to say ‘skip.’ Both the total time for naming each set and the number of errors and skipped words were recorded. The final score was the number of words that students correctly named per minute. Another measure available from school files was performance on the Illinois Standards Achievement Test of reading which students took when they were in the fourth grade.

Prior to the intervention, students completed questionnaires to obtain information about social relationships and individual characteristics. Notably, students were asked to nominate up to five of the quietest students in their class and up to five students who have the most to say during class discussions. Information about family background and language use was obtained from a parent questionnaire. Questionnaires in both Spanish and English were provided.

After the Wolf Unit, students completed an assessment battery providing extensive information about learning outcomes. Included in the battery were a 100-item sentence verification test that provided a broad, although not deep, assessment of concepts and information acquired from the Wolf Unit; a 50-minute individually-written essay in which students explained their own decisions about whether the pack of wolves should be eradicated; an individual oral interview about an analogue to the wolf question, whether whaling should be allowed; another transfer test, a 50-minute essay writing task about a moral and practical dilemma, whether a boy should tell on a classmate who cheated in a model car race; and a storytelling task.

The current paper only reports children’s performance on the storytelling task, which was included in the assessment battery to determine if students had made generalizable improvements in language proficiency (Larsen-Freeman, 2013). A research assistant asked individual students to tell a story prompted by the wordless picture book, *Frog, Where Are You?* (Mayer, 1969). The assistant elicited the story following the procedure described by Berman and Slobin (1994). First, students looked through the book so that they could get a sense of the whole story and prepare for the task. Then, they were asked to tell a story while they turned the pages of the book. Assistants used standardized prompts depending on students' behavior. For instance, assistants were to ask, "Can you tell more," if students stopped in the middle of the story. Students were allowed to code switch to Spanish if they did not know the English expression, although this rarely happened.

Coding oral narratives

Language production coding. Students' oral narratives were transcribed following the Systematic Analysis of Language Transcripts (SALT) conventions (Miller & Chapman, 2010). The transcripts were segmented into communication units (C-units). A C-unit is "a proposition or group of words that cannot be further broken down without loss of essential meaning" (Loban, 1976, p. 9). A C-unit represents an independent clause with all its modifiers, that is to say, one main clause plus its subordinate clauses. For example, "While the boy and the dog were sleeping, the frog decided to go out for a walk" is considered a C-unit, in which "the boy and the dog were sleeping" is a subordinate clause and "the frog decided to go out for a walk" is the main clause. Clauses with compound predicates were further segmented; for example, "The gopher popped out and bit the boy on his nose" was parsed into two C-units.

One analyst segmented all the transcripts into C-units. A different analyst independently segmented 20% of the transcripts. The percentage of agreement between the two analysts in C-unit segmentation was 98.4% (Cohen's κ= .89). Next the segmented transcripts were coded following SALT conventions. The codes included bound morphemes (marked by a slash; e.g., take/3s), mazes (in parenthesis), omissions (denoted by an asterisk), pauses (denoted by a colon), and errors (denoted by word-level error code [EW:word] and utterance-level error code [EU]). A word-level error was marked when a word was used incorrectly; e.g., the dog falled [EW:fell] out of the window. An utterance-level error was marked when the error was not simply associated with a certain word; e. g., then the boy up (uh the) the rock/s [EU]. A maze referred to false starts, repetition and revisions, filled pauses, and part words, e.g., the dog was (barking um um) barking at the beehive. Omissions included omitted words (e.g., the boy went *to the forest) and omitted bound morphemes (e.g., the boy look/*ed in the hole). Pauses included within-utterance pauses (denoted by a colon followed by the amount of time) and between-utterance pauses. A semi-colon followed by a colon was used to indicate pauses within the same speaker and two adjacent colons were used to indicate pauses between two speakers. Pauses were only noted when more than 3 seconds. The subordination index is represented by the SI-number, which refers to the total number of clauses in one C-unit. The following example includes bound morphemes, mazes, between- and within-utterance pauses, omissions and the code for subordination index. C stands for child and E refers to examiner.

> C (Once there) :04 once there was a boy (who got a fr*) who had a frog and a dog [SI-2].
> C But now we/'re gonna tell you (how the story) how they got the frog [SI-2].
> E Ok.
> : : 05 *between speaker pause*
> C (O*) one day the frog got lost [SI-1].
> ; : *04* *between utterance pause*
> C Then : while : the (k*) boy was sleep/ing (um um) the boy woke up the next day (find*) find/ing out that (the) the frog *was miss/ing [SI-3].

The 13 measures obtained from SALT representing aspects of language production were factor analyzed using maximum likelihood estimation and varimax rotation. The measures loaded on five clearly defined language factors, as is shown in Table 2. The five factors were language volume *or length* (total number of C-units, total number of complete words, and total number of different words), *syntactic complexity* (mean length of utterance in words and in morphemes, and subordination index), *verbal fluency* (words per minutes, between and within utterance pauses), *mazes* (number of mazes, maze words, percentage of maze words, and utterance with mazes), and *errors and omissions* (omitted words, omitted bound morphemes, word-level errors, and utterance-level errors).

[Insert Table 2 here]

Story element coding. Story elements were defined on the basis of Stein and Glenn’s (1979) story schema theory. The process of storytelling is “a product of interaction between incoming information and strategies, mental operations, and structures inherent in the [storyteller]” (Stein & Glenn, 1979, p. 54). The fundamental elements in a story include a setting and an episode system. The episode system (i.e., a collection of different episodes) can be further split into several components—initiating event, internal response, external response, and consequence. The episode structure of the frog story is presented in Figure 1 and the corresponding coding system is described in Figure 2. The first author coded all the transcripts and a different coder coded 20% of the data to check the reliability of each story element.

[Insert Figure 1 and 2 here]

As shown in Figure 2, a story includes *critical events* and *minor events*. Critical events are activities that advance the main theme of the story, which consist of the boy’s explicit attempts to search for the frog in different places. Coding reliability for each critical event is given below. Minor events refer to the activities of supporting characters (i.e., bees, gopher, owl, and deer), which are not critical to the plot development but are mentioned in response to the picture book; for example, the dog was chased by bees or the boy was bitten by a gopher. The intercoder percentage of agreement for *minor events* was 93.5% (Cohen's κ = .83).

Setting refers to the introduction of main characters and the description of the temporal and physical context in which the story takes place. In the frog story, the three main characters are the boy, dog and frog. The story began in the boy’s bedroom during nighttime. The intercoder percentage of agreement for *setting* was 99.6% (Cohen's κ= .96).

Initiating Event refers to the event that initiates the response of a main character. In the frog story, it is that the frog is discovered to be missing. When the boy and the dog woke up in the morning and found that the frog had escaped from the jar, they started a sequence of actions to try to find the frog. The intercoder percentage of agreement for *initiating event* was 98.7% (Cohen's κ= .87).

Internal Response refers to "the psychological state of a character after an event" (Stein & Glenn, 1979, p. 65) and takes three forms—*goals, cognitions,* and *affects*. *Goals* stands for "desires or intentions of a character" (Stein & Glenn, 1979, p.65), i.e., the purpose for taking an action. In the frog story, for example, to find the frog is the most important goal set by the boy at the beginning of the story. *Cognitions* are characters' thoughts or beliefs. The search for the frog is guided by the main character's thoughts. *Affects* represent characters' emotional reactions such as happiness, sadness, anger, worry, or fear. For example, when the boy couldn't find the frog in the places that he checked, he felt upset and worried. The intercoder percentage of agreement for *internal response* was 86.8% (Cohen's κ= .74).

External Response refers to a sequence of behavior that reveals the main character's attempts to change the "disequilibrium that was caused by the initiating event" (Stein & Glenn, 1979, p. 67). External response includes external events with or without a purpose. Events with an explicitly specified purpose are *attempts*, whereas events that do not explicitly specify a purpose are *actions*. For example, the event "the boy called for his frog out of the window" is an attempt because the boy's goal was to find the frog. However, the event "the boy went to the forest" is an action because no purpose was specified in this statement. The intercoder percentage of agreement for *attempts* was 97.5% (Cohen's κ= .86) and for *actions* was 94.0% (Cohen's κ= .82).

Consequences in the episode system describe whether the main character achieves a goal or not. A consequence could be an *outcome* of an attempt to reach a goal or an *end state*. For example, the boy looked into a hole in the ground but he did not find the frog there; thus, the *outcome* to the attempt is the failure to find the frog. The intercoder percentage of agreement for outcome was 86.3% (Cohen's κ= .77). The main character stops making further attempts either by achieving the goal or completely giving up. In the frog story, the boy and the dog eventually found the frog and took a little frog back home. This consequence is the *end state* or the concluding statement of the story. The intercoder percentage of agreement for *end state* was 98.5% (Cohen's κ= .85).

Coding for multi-link reasoning. In narrative text, multi-links are identified as interconnected sequential events. Each event describes the attempt to achieve the goal of the main character. The failure of one attempt causes the next attempt. In other words, the occurrence of subsequent events is based on the outcome of previous events. By linking sequential events together, a causal reasoning chain is formed, which makes the story more organized and coherent. The coding of multi-link reasoning chains is based on the identification of attempts and outcomes. Table 3 lists the attempts and outcomes in the *Frog, Where Are You* story. The outcome to an attempt refers to the failure (fail to find the frog) or the success (find the frog) of an attempt. An example of multi-link reasoning structure is also presented in Table 3.

[Insert Table 3 here]

The main character keeps making attempts until the fulfillment of the goal. Every event in this causal chain is considered as a step in multi-link reasoning process. The intercoder percentage of agreement for *multi-link reasoning chains* was 93.3% (Cohen's κ= .83).The total number of links is calculated and compared between the different intervention conditions.

Results

Initial language proficiency

Two pretests were administered to assess students' reading level and basic oral English proficiency: Gates-MacGinitie reading comprehension and rapid naming speed. Students received the Illinois Standards Achievement Test (ISAT) six months before the intervention when they were in the fourth grade. However, 24% of the participants did not take the pre-intervention ISAT due to high student mobility in the cooperating schools. Table 4 summarizes the descriptive statistics on the two pretests and the fourth-grade ISAT reading test. Separate ANOVAs were conducted in which the two pretests and normed ISAT reading scores were dependent variables, condition was a fixed effect, and classroom was a random effect to account for variance due to the teacher or the student cohort. The results indicated no condition difference for pretest reading comprehension, F (2, 192) = 0.64, p = .53, the rapid automatized naming test of oral English proficiency, F (2, 192) = 0.58, p = .56, or fourth-grade ISAT reading, F (2, 142) = 0.69, p = .51.

[Insert Table 4 here]

Language production

Table 5 presents standardized factor scores (percentaged z-scores, Strauss, 1980, with M=50, SD=20) by condition for each of the five language production factors. The overall difference among the three conditions was examined in a MANCOVA with the five language factor scores as dependent variables. Age of English acquisition (English in the first grade), parents' education level and intervention condition were entered as fixed effects. Reading comprehension and oral English proficiency (class-mean-centered scores) and talkativeness (peer nominations of talkativeness minus nominations for quietness divided by the number of students in the class) were covariates. Class-mean reading comprehension and oral English proficiency were entered to account for between-classroom variance. Four students with an Individualized Education Plan (IEP) were excluded from the analysis. An outlier in fluency was also dropped. The student produced a short story (235 words) during a relatively long time (6 minutes) and paused frequently. After these exclusions, 205 subjects remained in the language production analysis.

[Insert Table 5 here]

There was a significant overall condition effect on language production, Wilks' Lambda = .89, F (10, 380) =2.34, p = .011, and a significant difference between students who used and did not use English in the first grade, Wilks' Lambda = .92, F (5, 190) = 3.16, p < .01. Individual-level reading, individual-level oral English proficiency, class-level reading, and class-level oral English proficiency were all significant predictors, ps < .001. Parent education level and student talkativeness did not significantly predict overall language production.

Since class-mean pretest scores were significant in the multivariate analysis, we built two-level models for univariate analyses of each language factor with classroom (n=18) as the second-level factor and class-mean reading comprehension and oral English proficiency as the predictors for the random effect. All the other covariates were entered at the first (individual) level. The level-1 equation was

$$(\text{Language factor})_{ij} = \beta_{0j} + \beta_{1j}\,(\text{condition}) + \beta_{2j}\,(\text{language}) + \beta_{3j}\,(\text{education}) + \beta_{4j}\,(\text{class-mean-centered reading}) + \beta_{5j}\,(\text{class-mean-centered oral English proficiency}) + \beta_{6j}\,(\text{talkativeness}) + e_{ij}$$

And the level-2 equation for random intercept was

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\,(\text{class-mean oral English proficiency}) + \gamma_{02}\,(\text{class-mean reading}) + e_{0j}$$

Language volume. There was no significant condition effect, F (2, 181) = 0.83, p = .44. However, there was a marginal age of acquisition of English effect favoring students who used English in the first grade, F (1, 181) = 3.53, p = .062, standardized effect size δ = .30. No other predictors were significant and no classroom difference was observed.

Syntactic complexity. There was a significant condition effect, F (2, 181) = 4.12, p = .018. CG students outperformed control students, mean difference M_{diff} = 0.46, F (1, 181) = 8.25, p < .01, δ = .50. CG students' narratives had more complicated syntax than DI students' narratives; however, the difference was not significant, F (1, 181) = 1.89, p = .17, nor was the difference between DI and control students, F (1, 181) = 2.09, p = .15. There was a marginal age of acquisition of English effect favoring students who spoke English in the first grade, F (2, 181) = 2.93, p = .089, δ = .28. Individual-level reading comprehension and oral English proficiency significantly predicted syntactic complexity, ps < .02. No other predictors were significant and no difference among classrooms was observed.

Language fluency. There was no significant condition effect, F (2, 181) = 0.20, p = .82. However, age of acquisition of English showed a significant effect favoring students who used English rather than Spanish or English-Spanish mixture in the first grade, F (1, 181) = 4.88, p = .028, δ = .36. Individual-level oral English proficiency was a significant predictor of language fluency, F (1, 181) = 4.31, p = .039. The results also indicated a teacher/cohort effect; about 8% (intra-class correlation; ICC) of the variance was accounted for differences among classrooms. No other predictors were significant.

Mazes. A random slope for individual-level reading score was included in the model rather than adding classroom-level reading to the random intercept. The reason was that individual reading comprehension was positively correlated with maze production in classrooms with low average English proficiency, but negatively correlated in classes with high average English proficiency. After allowing the relationship between individual reading ability and maze production to vary between classrooms, there was a nearly significant main effect of condition, F (2, 164) = 3.02, p = .052, and a significant age of acquisition of English effect showing that students who spoke Spanish or a mixture of Spanish and English in the first grade generated more mazes, F (1, 164) = 3.92, p = .049, δ = .32. Post hoc comparisons indicated that CG students produced more mazes than control students, t (164) = 2.21, p = .029, δ = .49, and DI students also produced more mazes than control students, t (164) = 2.04, p = .043, δ = .47. There was no significant difference between CG and DI condition, t (164) = .10, p = .92. Both individual-level oral English proficiency and reading comprehension were found to be significant predictors of mazes, ps < .05. There was a teacher/cohort effect in maze production; about 9% (intra-class correlation; ICC) of the variance was accounted for by differences among classrooms.

Errors and omissions. There was a significant condition effect, F (2, 181) = 4.37, p = .014. The contrast between CG and control group was significant, mean = -0.32, t (181) = -2.11, p = .036, δ = .39, as well as the contrast between DI and control group, mean = -0.43, t (181) = -2.84, p < .01, δ = .53, which indicated that both CG and DI students generated fewer errors than control students. There was no significant difference between CG and DI condition, mean = 0.11, t (181) = 0.72, p = .47. Age of acquisition of English did not show a significant effect. Students in classes with lower average reading comprehension generated more errors, F (1, 181) = 8.16, p < .01 and poor readers within classes had more errors and omissions in their storytelling, F (1, 181) = 3.50, p = .063.

Story elements

As a measure of communicative competence, an aggregate story quality score, called Essential Story Elements, was defined as the sum of descriptions in six critical story schema categories, namely *setting, initiating event, internal response, attempt, outcome,* and *end state*. Essential Story Elements advance the main theme of the story, describing story characters' goals, feelings, thoughts, attempts, and outcomes in relation to central story events; whereas extended descriptions of minor events, or actions that do not signal purpose, are indicative of stories of lower quality. The aggregate Essential Story Elements score and number of non-repetitive C-units in each story schema category are presented in Table 5.

Factors influencing Essential Story Elements were evaluated by conducting a two-level regression analysis with classroom as the second-level factor and class-mean centered reading comprehension, oral English proficiency, and talkativeness as covariates for the random effect. Intervention condition, age of acquisition of English, and parent education level were entered as fixed effects. The results indicated a significant overall condition effect on Essential Story Elements, F (2, 180) = 3.18, p = .044. CG students outperformed DI students, mean difference M_{diff} = 2.49, t (180) = 2.04, p = .042, δ = .38. CG students also outperformed control students, M_{diff} = 2.80, t (180) = 2.31, p = .022, δ = .43. However, there was no significant difference between DI and control students, t (180) = 0.25, p = .80. A marginal difference was observed between students who did or did not use English in the first grade, F (1, 180) = 2.89, p = .091, δ = .28. The between-classrooms effect was not significant. Neither class-level nor individual-level reading nor oral English were significant. Follow up tests of individual story elements, involving either Poisson regression analysis or negative binomial regression analysis, depending on the distribution of the counts of story elements, indicated that CG students significantly exceeded students in the other two conditions in elaborations of *initiating events* and *outcomes*.

Students who told longer stories may have had a greater chance to produce Essential Story Elements, but it is also possible that long stories contained extensive descriptions of minor events. We reanalyzed Essential Story Elements including the length factor from the language production analysis as a covariate. After controlling for length, the condition difference remained significant and the effect was larger, F (2, 179) = 7.05, p < .01. CG students outperformed DI students, M_{diff} = 2.14, t (179) = 2.31, p = .022, δ = .45, and control students, M_{diff} = 3.44, t (179) = 3.72, p < .01, δ = .72. No difference between students who did or did not use English in the first grade was observed, F (1, 179) = 0.32, p = .57, but parents' education level showed a significant effect, F (2, 179) = 3.25, p = .041. The higher parents' education level, the more Essential Story Elements students produced. Class-mean reading comprehension was a significant predictor, F (1, 179) = 4.70, p = .031. If a child came from a class with higher average reading scores, he or she is likely to produce more Essential Story Elements. Other covariates were not significant.

No condition difference was found in description of minor events, F (2, 194) = 0.50, p = .61, and there was no effect of age of acquisition of English, F (1, 194) = 0.03, p = .86. Students with higher oral English proficiency generated more minor events, F (1, 194) = 4.82, p = .029.

Multi-link reasoning

A multinomial logistic regression analysis was conducted to examine whether there was an intervention effect on the length of causal chains in the children's frog stories. The covariates in this analysis were age of acquisition of English, reading comprehension, basic oral English proficiency, talkativeness, and length of narrative. Number of links in the longest causal chain in a student narrative was treated as the outcome variable. There were four categories: 1 link (N = 114), 2 links (N = 47), 3-4 links (N = 33), and 5-7 links (N = 11). We used 2 links as the reference category for number of links in chains and the control group as the reference category for intervention condition. The results indicated a significant condition effect, χ^2 (6, N=205) = 15.00, p = .020.

As shown in Table 6, CG students had a higher probability of generating chains with 5-7 links (odds ratio CG/control = 15.18 vs. DI/control = 6.36), whereas DI students had a higher probability of generating 1-2 link chains (odds ratio CG/control = 1.30 vs. DI/control = 2.67). The two conditions were comparable in generating 3-4 link chains (odds ratio CG/control = 1.02 vs. DI/control = 0.82). We calculated the predicted probability of generating multi-link chains of different lengths for each condition which are presented in Figure 3. The figure shows that CG students were more likely to connect several story events into causal chains while DI and control students tended to describe each story event separately.

[Insert Table 6 here]

[Insert Figure 3 here]

Classroom talk during the Wolf Unit

Figure 4 shows the rate per minute of any of the 76 academic vocabulary words by teachers and students, respectively, in the twelve CG and twelve DI classrooms. The classrooms are ordered from the lowest rate to the highest rate of use within condition. The expected differences between CG and DI classrooms are readily apparent. Among students, the rate per minute of academic vocabulary was over twice as high in CG classrooms (M_{CG} = 2.34, SD = 1.28) as compared to DI classrooms (M_{DI} =1.01, SD = 0.72). Among teachers, the rate of use was four times higher among DI teachers (M_{DI} = 1.44 SD = 0.68) than CG teachers (M_{CG} = 0.33, SD = 0.25). We built a two-level Poisson model to compare condition differences in the use of academic vocabulary, with classroom as the second level factor, four-minute episodes within classrooms as first-level factor, and duration of talk as a first-level covariate. The results indicated that, after controlling for duration of talk, students in the CG condition produced significantly more academic vocabulary words than students in the DI condition, F (1, 122) = 16.56, p <.001, and teachers in DI condition produced significantly more academic vocabulary words than teachers in the CG condition, F

(1, 122) = 33.36, p <.001. In CG classrooms, students in aggregate produced 86% of the academic vocabulary while 14% was produced by teachers. In contrast, in DI classrooms teachers produced 60% of the academic vocabulary whereas the students in aggregate produced 40%. The foregoing analysis was completed with all CG and DI classrooms [N=24]. When the sample is restricted to classrooms with a high proportion of Hispanic students [N=12], the results are essentially the same.

[Insert Figures 4 here]

We searched the frog stories for the 76 academic vocabulary words. Only four students used a total of just three of the words. One CG student twice said *habitat*. Another CG student said *wander*. A DI student and a Control student used *howl*. The low rate of occurrence of this set of academic vocabulary words is not surprising since the words are not useful for talking about the adventures of a boy, a dog, and a frog.

To give an impression of discourse features of classroom talk we selected two short excepts, one from a DI classroom and one from a CG classroom, that we judge to be representative of the discourse in the DI and CG classrooms. The two selected classes represent mid-level performance in terms of rate of use of academic vocabulary.

The first excerpt is from a direct instruction classroom. The teacher comments on a student's idea and asks everyone to evaluate its plausibility. Some students support the idea, although in face of the strong implication from the teacher suggesting the opposite, they abandon their turns before they express themselves completely. The teacher explains her thinking and guides students to follow her. The contrary idea is never mentioned again by the students or the teacher. This is an example of thoughtful dialogue inasmuch as the teacher picks up on and responds to student ideas about a fundamental issue in the Wolf Unit.

Teacher	Okay. That is a possibility but based on the facts uh the facts and the research that we've been doing, is that likely?

Student 1 Yes.

Student 2 Unless you do something, it's not gonna…

Teacher Is it?

Student 3 It's outrageous.

Student 4 Yeah because they said (in the um, the um) one of those things about (wolves) wolves. . .

Teacher The people are afraid of the wolves. That is a fact. They (they) do think they're dangerous, but based on the research and the numbers that we looked at, are they really, truly a threat? There've been instances but are those instances, do they occur frequently or not frequently?

Students No. Uh-uh. Not frequently.

The following except is from a collaborative group work classroom. Student 1 states a position and is immediately challenged by Student 2. Student 1 then provides a reason to support his position. His idea is picked up by Student 3 and further developed into a chain of reasoning. Student 4 supports the emerging idea. Student 5 starts to express the counter-position, but before she can finish, Student 4 asks a challenging, “Why?” Student 5 provides a reason to support her viewpoint. The excerpt illustrates a frequent pattern in CG classrooms in which students challenge each other, ask for elaboration and explanation, express complicated ideas, build on each other’s thinking, and co-construct ideas.

Student 1 I think that we shouldn't.

Student 2 Why?

Student 1 Because you just hurting the, population of (of) the wolves.

Student 3 I say they shouldn't because, um, the wolves kill the elk, and when the elks are dead, there are more trees, so there's more oxygen for us.

Student 4 I was going to say that.

Student 3 So the wolves make it better for us.

Student 5 Hm. I- I think they should hire people to…

Student 4 Why?

Student 5 Becaaause, the wolves are just killing all the animals.

Discussion

The major finding of this study is that fifth grade English language learners who participated in collaborative groups during a six-week unit on wolf reintroduction and management told more elaborated and cohesive stories than comparable students who received direct instruction or were wait-listed controls. As compared to students in the other two conditions, students who had interacted in collaborative groups constructed stories that contained significantly more explanation of essential story elements and generated significantly longer causal reasoning chains connecting story elements. The concepts and vocabulary needed to tell a story from the wordless picture book, *Frog, Where Are You,* which served as the story prompt, bear little relationship to the specific concepts and vocabulary taught in the Wolf Unit. Thus, the superior performance of students who participated in collaborative groups implies that they acquired, or further developed, some generalized competencies in language and thought.

With respect to basic features of language production, the overall difference between instructional conditions was significant, largely because of the contrast between the CG and control students with DI students in the middle on most measures. As compared to control students, CG students produced stories with more complicated syntax. Both CG and DI students made fewer omissions and word- and utterance-level errors than control students. But, after allowing the relationship between individual reading ability and mazes to vary from class to class, both CG and DI students were found to generate more mazes than control students. There were no significant differences between CG and DI students on any of the basic language production measures.

CG students outperformed DI and control students in the production of Essential Story Elements with or without controlling for story length. The better performance of CG students in Essential Story Elements suggests that students who had six-weeks of collaborative group work focused more on the critical events and the main theme of the story, which implies that CG students had developed a better understanding of story construction and improved communicative competence.

Regarding the development of multi-link causal reasoning, CG students made more connections between story events than DI or control students. CG students had a significantly higher probability of generating chains with 5-7 links whereas DI students had a significantly higher probability of generating 1-2 link chains, as compared to control students. The intervention conditions did not differ in the likelihood of chains with 3-4 links.

The likely reason for the generally superior performance of CG students is that collaborative discussion provides more opportunities for high quality student talk than the teacher-dominated discourse prevalent during direct instruction. An indicator of quality talk is use of academic language. In present study CG students had twice as high a rate as DI students in use of academic vocabulary words.

In CG discussions, each student acts as a provider as well as a receiver of information. One person's statement is likely to be extended or evaluated by other students. If the statement is not clear, clarification may be requested. If students agree on the same idea, supplementary evidence may be offered to support the agreed upon point of view. If students disagree with one another, counterarguments and rebuttals will be made to support different opinions. One student's talk is usually extended by other students through making connections between their own opinions and the other student's ideas. In comparison, DI students are less likely to extend the talk of teachers or peers. Usually students in DI classrooms are only receivers of information. They have few opportunities to initiate ideas. Teachers do much of the talking and students just answer questions, usually with answers that are brief and unelaborated. The thinking required for extended talk may be suppressed in teacher dominated lessons (Nystrand & Gamoran, 1991).

Probably, the major reason that students in the CG condition generated more cohesive stories and longer multi-link reasoning chains than students in the other two conditions is that collaborative group work provided more opportunities to use language to make connections. Morris and his colleagues (2013) examined the frequency of use of the coordinating conjunctions *because, so, if, then, and*, and *but*, which are low-inference indicators of connected talk and relational thinking (Lin et al., 2015). Students' rate of use of coordinating conjunctions was four times higher in CG than DI classrooms enrolled in the present study. Thus, it is highly plausible that CG students generated more elaborated and connected stories because of the experience of expressing elaborated and connected ideas during collaborative group work. DI students, in contrast, depended on teachers to initiate ideas and make connections and the students were left with only small pieces to add to a narrative largely told by the teacher.

CG students are encouraged to elaborate not only *what* but also *how* and *why* (Clark et al., 2003). CG students are allowed to freely express ideas and are expected to provide supporting reasons and evidence during discussions. Such experience stimulates students to generate more convincing arguments and fuels the development of multi-link reasoning and other forms of relational thinking (Lin et al., 2012; Reznitskaya et al., 2009). A well-structured story has logically arranged event sequences based on explicit statements of the relationships between events. With the elaboration of goals and outcomes, students are more capable of creating causal links between events. "Understanding why and under what circumstances people act on beliefs, true or false, may well be part of a more general scheme of understanding what causes events, states, state changes, and actions" (Trabasso, Stein, Rodkin, Park Munger, & Baughn, 1992, p. 164). The appreciation that one must strive to explain the causal relationships between events, rather than simply say what happened, is probably foundational for the development of multi-link reasoning ability.

Both the Essential Story Elements measure and the multi-link reasoning measure, developed for the analyses reported in the current paper, contribute to the methodology for studying children's narratives. Previous studies have relied on the Narrative Scoring Scheme (Miller et al., 2006) as an index of children's ability to produce well-formed narratives as defined in story schema theory. In this scheme, judges rate the quality of stories in terms of each story schema category. The overall score is the sum of the ratings in the various categories. The trouble is that the scale is subjective and different raters may apply a different standard. A given rater's standard may drift over the course of rating a number of stories. Ratings are subject to halo effects from the rater's general impression of stories. The Essential Story Elements measure employed in this study is also based on story schema theory (Stein & Glenn, 1979), but each C-unit was judged in terms of whether it expressed one of the story schema categories. This is a big improvement over ratings on an arbitrary and subjective scale, since it entails low-inference rule-governed assignment to categories. When evaluating the stories of children with limited English, Essential Story Elements is probably less vulnerable to negative halo effects than the Narrative Scoring Scheme; the poor general impression created by disfluencies may induce lower ratings of the stories of ELLs.

The multi-link reasoning analysis is an extension of the casual network model created by Trabasso and his associates (Trabasso & van den Broek, 1985; Trabasso, van den Broek, & Suh, 1989). It builds on the idea that a good story contains causally-related events and well-structured story episodes. However, Trabasso and colleagues used the causal model to analyze the stories of adult authors, not the stories told by children. So, it is a big step forward to analyze child-created stories in terms of causal connectedness.

Multi-link reasoning requires the ability to identify and make connections between events. More developed multi-link reasoning ability is reflected by generating longer causal chains. Lin and her associates (2011) found that CG students developed multi-link reasoning skills during the Wolf Unit which they applied when writing policy decision letters presenting arguments about the wolf question. Longer causal chains were identified in CG students' letters than DI students' letters. In this study, we documented that longer multi-link reasoning chains appear not only in student-produced argumentative text, but also appear in their narrative text. Consistent with Lin et al.'s study, collaborative group work was found to have a positive effect on students' ability and disposition to link events in causal chains.

Several characteristics of children affected features of their stories. Students who acquired English early [English only in the first grade versus Spanish only or a mixture of Spanish and English in the first grade] produced longer stories, showed higher verbal fluency, and had fewer repetitions and revisions.

Birdsong (2005) maintained that the chances for native-like attainment of a second language decrease with age of acquisition. In this study, age of acquisition of English had more effect on the language production measures than on the story quality measure or multi-link reasoning. It is likely that age of acquisition has a stronger effect on basic linguistic proficiency than communicative competence or reasoning. Children are still capable of developing communicative competence and thinking skills at a high-level even when they learn a language at an older age.

A limitation of this study is that we did not give a pretest measure of storytelling, or any other pretest measure of oral language production, beyond the level of discrete words assessed by asking children to rapidly name common objects. A pretest measure of oral discourse production no doubt would have explained additional variance in individual oral narrative ability and enabled more sensitive tests of intervention effects. Another limitation is that, although the study was fairly large compared to most previous intervention studies with ELLs, 18 classrooms is at the lower margin for fitting multi-level models that account for teacher and cohort effects. Classroom-level predictors explained some variance in language production measures but effects at the classroom level were weakly estimated.

Overall, collaborative group work improved Spanish-speaking ELL's oral narrative skills. They were more willing and able to express complicated ideas, as indicated by the greater syntactic complexity of their utterances. More attempted speech was produced by low proficiency students who participated in collaborative group work, although this was associated with more frequent mazes. Stories produced by students who had interacted in collaborative groups were more coherent and causally structured. They were more likely to elaborate essential story elements and organize events into causal chains. All these gains suggest that collaborative group work is a promising approach to promote ELL's language proficiency and, at the same time, their cognitive development.

For children who are second language learners and cannot always express themselves well in their new language, it is natural to conclude that they are facing language difficulties rather than having thinking problems. However, we cannot assume that children's natural ability to think will come out as soon as they know enough words and have control of the grammar of the second language. Abundant time is spent in trying to improve basic language skills of second language learners while less attention is given to thinking or to understanding science, social science, arts, and humanities concepts (Moll, 2010). Meanwhile, native speakers are getting the chance to improve their thinking and conceptual understanding as well as their language. Bilingual educators should recognize the synergy that comes from the co-evolution of elemental language skills, communicative competence, thinking, and conceptual understanding.

REFERENCES

Anderson, R. C., Chinn, C., Waggoner, M., & Nguyen-Jahiel, 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.-Y., Reznitskaya, A., … Gilbert, L. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction, 19*, 1-46.

Arreaga-Mayer, C., & Perdomo-Rivera, C. (1996). Ecobehavioral analysis of instruction for at-risk language minority students. *Elementary School Journal, 96*(3), 245-258.

Avila, E., & Sadoski, M. (1996). Exploring new applications of the keyword method to acquire English vocabulary. *Language Learning*, *46*(3), 379-395.

August, D., & Shanahan, T. (2008). *Developing reading and writing in second-language learners: Lessons from the Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

Baker, C. (2011). *Foundations of bilingual education and bilingualism (5th ed.).* Clevedon, UK: Multilingual Matters.

Bandura, A. (1986) *Social foundations of thought and action: A social-cognitive view*. Englewood Cliffs, NJ: Prentice-Hall.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Knoll and A. M. B. de Groot (Eds), *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press. (pp. 109-127)

Candelaria, M. A., & Llorente, A. M. (2009). The assessment of the Hispanic child. In C. Reynolds, & E. Fletcher-Janzen (Eds.), *Handbook of Clinical Child Neuropsychology* (pp. 401-424). Springer US. doi: 10.1007/978-0-387-78867-8_18

Cazden, C. (2011). Dell Hymes's construct of "Communicative Competence." *Anthropology & Education Quarterly*, 42, 364-369.

Chinn, C. A., Anderson, R. C., & Waggoner, M. A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly*, *36*(4), 378-411.

Clark, A., Anderson, R. C., Kuo, L., Kim, I. H., Archodidou, A., & Nguyen-Jahiel. K. (2003). Collaborative reasoning: Expanding ways for children to talk and think in school. *Educational Psychology Review, 15*, 181-198.

Common Core State Standards Initiative. (2010). *Common Core State Standards for English Language arts & literacy in history/social studies, science, and technical subjects Appendix A*. Retrieved from http://www.corestandards.org/assets/Appendix_A.pdf

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review*, 56, 18-36.

Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English language learners in US schools: An overview of research findings. *Journal of Education for Students Placed at Risk*, *10*(4), 363-385.

Goldenberg, C. (1992). Instructional conversations: Promoting comprehension through discussion. *The Reading Teacher*, *46*(4), 316-326.

Goldenberg, C. (1996). *Latin American immigration and U.S. schools*. Social Policy Reports: Society for Research in Child Development, 10(1).

Hymes, D. (1972). On communicative competence. In J. B. Pride and J. Homes (Eds.). *Sociolinguistics: Selected readings*. (pp. 269-293) Harmondsworth, UK: Penguin Books

Jadallah, M., Miller, B., Anderson, R. C., Nguyen-Jahiel, K., Archodidou, A., Zhang, J., & Grabow, K. (2009). Collaborative Reasoning about a science and public policy issue. In M. McKeown and L. Kucan (Eds.) *Bringing reading research to life: Essays in honor of Isabel L. Beck.* New York: Guilford Press. (pp. 170-193)

Johnson, D. W., & Johnson, R. T. (2009). Energizing learning: The instructional power of conflict. *Educational Researcher*, 38, 37-51.

Kamps, D., Abbott, M., Greenwood, C., Arreaga-Mayer, C., Wills, H., Longstaff, J., ... & Walton, C. (2007). Use of evidence-based, small-group reading instruction for English language learners in elementary grades: Secondary-tier intervention. *Learning Disability Quarterly*, 153-168.

Kelcey, B., & Carlisle, J. F. (2013). Learning about teachers' literacy instruction from classroom observation. *Reading Research Quarterly*, 48, 301-317.

Kindler, A. L. (2002). *Survey of the states' limited English proficient students and available educational programs and services: 2000-2001 summary report*. Washington. D. C: National Clearinghouse for English Language Acquisition.

Larsen-Freeman, D. (2013). Transfer of learning transformed. *Language Learning*, 63, 107-129.

Lin, T.-J., Anderson, R. C., Hummel, J. E., Jadallah, M., & Miller, B. W., Nguyen-Jahiel, K., Morris, J. A., Kuo, L. J., Kim, I.L., Wu, X., & Dong, T. (2012). Children's use of analogy during Collaborative Reasoning. *Child Development, 83,* 1429-1443.

Lin, T.-J., Anderson, R. C., Jadallah, M., Nguyen-Jahiel, K., Kim, I.-H., Kuo, L.-J., Miller, B. W., Dong, T., Wu, X., & Li, Y. (2015). Social influences on the development of relational thinking during small-group discussions. *Contemporary Educational Psychology*, 41, 83-97.

Lin, T.-J., Ma, S., Zhang, J., Nguyen-Jahiel, K., Anderson, R. C., Morris, J. A., … Jadallah, M. (April, 2011). *Nurturing conceptual understanding and systems thinking.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L.G. (2000). *Gates-MacGinitie Reading Tests (4th ed.), Level 4, Form S*. Itasca, IL: Riverside Publishing.

Magliano, J. P. (1999). Revealing inference process during text comprehension. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 55-76). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mayer, M. (1969). *Frog, where are you?* New York: Dial Books for Young Readers.

McCaslin, M., Good, T. L., Nichols, S., Zhang, J., Wiley, C. R., Bozack, A. R., … Cuizon-Garcia, R. (2006). Comprehensive school reform: An observational study of teaching in grades 3 through 5. *Elementary School Journal*, *106*(4), 313-331.

Miller, J. F., & Chapman, R. S. (2010). Systematic Analysis of Language Transcripts [computer software]. Madison, WI: Language Analysis Laboratory, Waisman Center, University of Wisconsin.

Miller, J., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research & Practice, 21*, 30–43.

Moll, L. C. (2010). Mobilizing culture, language, and educational practices: Fulfilling the promises of Mendez and Brown. *Educational Researcher*, 39, 451-460.

Morris, J., Miller, B., Anderson, R. C., Lin, T.-J., Nguyen-Jahiel, K., Sun., J., … Wu., X. (2013). *Instructional discourse and argumentative writing*. Champaign, IL: Center for the Study of Reading, University of Illinois

Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' high-level comprehension of text: A meta-analysis. *Journal of Educational Psychology, 101*, 740-764.

Nystrand, M., & Gamoran, A. (1991). Instructional discourse, students' engagement, and literature achievement. *Research in the Teaching of English, 25*, 261-290.

Padolsky, D. (2011). *How many school-aged English language learners (ELLs) are there in the U.S?* NCELA FAQ No. 1. Retrieved September 15, 2012, from http://www.ncela.gwu.edu/faqs/view/4.htm

Piaget, J. (1976/1947). *The psychology of intelligence*. New York: Littlefield.

Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Clevedon, UK: Multilingual Matters.

Pressley, M., & el Dinary, P. B. (1997). What we know about translating comprehension-strategies instruction into practice. *Journal of Learning Disabilities*, 30, 486-488

Reznitskaya, A., Kuo, L., Clark, A., Miller, B., Jadallah, M., Anderson, R. C., & Nguyen-Jahiel, K. (2009). Collaborative reasoning: A dialogic approach to group discussions. *Cambridge Journal of Education, 39*, 29-48. doi:10.1080/03057640802701952

Rojas-Drummond, S., & Mercer, N. (2003). Scaffolding the development of effective collaboration and learning. *International Journal of Educational Research*, *39*(1), 99-111.

Saunders, W. M., & Goldenberg, C. (2007). The effects of instructional conversation on transition students' concepts of friendship and story comprehension. In Horowitz, R. (Ed.), *Talking Texts: How Speech and Writing Interact in School Learning.* Hillsdale, NJ: Erlbaum Associates.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 174-215.

Snow, C. E. (2014). Input to interaction to instruction: Three key shifts in the history of child language research. *Journal of Child Language, 41, Supplement S1*, 117-123.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Stein, M., Carnine, D., & Dixon, R. (1998). Direct instruction integrating curriculum design and effective teaching practice. *Intervention in School and Clinic*, *33*(4), 227-233.

Stein, N., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.). *New directions in discourse processing* (pp.53-120). Norwood, NJ: Ablex.

Straus, M. A. (1980). *The ZP scale: A percentaged Z score*. Durham, NH: Family Research Laboratory, University of New Hampshire.

Tharp, R., & Gallimore, R. (1989). *Rousing minds to life*. New York: Cambridge University Press.

Trabasso, T., Stein, N. L., Rodkin, P. C., Park Munger, M., & Baughn, C. R. (1992). Knowledge of goals and plans in the on-line narration of events. *Cognitive Development, 7*(2), 133-170.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*(5), 612-630.

Trabasso, T., van den Broek, P., & Suh, S. Y. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, *12*(1), 1-25.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard.

Webb, N. M., & Mastergeorge, A. M. (2000). The development of students' helping behavior and learning in peer-directed groups. *Cognition and Instruction*, 21, 361-428.

Wells, G., & Arauz, R. M. (2006). Dialogue in the classroom. *Journal of the Learning Sciences, 15(*3), 379-428.

Zhang, J., Anderson, R. C., & Nguyen-Jahiel, K. (2013). Language-rich discussions for English language learners. *International Journal of Educational Research*, 58, 44-60.

Table 1

Classroom demographic characteristics, language use, and pre-intervention test performance

Wave	Class	Condition[1]	Program[2]	Overall Class Size	Percent Hispanic in class (%)	Gender (M; F)	Age	Percent who speak Spanish with parents[3] (%)	Percent who used Spanish in first grade[3] (%)	Grade 4 Reading[4] M(SD)	Reading Comprehension[5] M(SD)	Basic English[6] M(SD)
1	1	CG	M	27	59	(15; 12)	10.8	75	69	207.4 (22.5)	13.4 (7.4)	57.0 (11.8)
	2	DI	M	28	64	(16; 12)	10.8	83	78	207.0 (24.8)	11.1 (8.4)	54.7 (11.0)
	3	Control	M	28	71	(17; 11)	10.7	75	80	204.3 (18.5)	14.1 (8.4)	54.7 (10.4)
	4	CG	B	23	100	(8; 15)	10.6	100	96	186.9 (19.9)	9.5 (6.5)	42.6 (10.6)
	5	DI	B	24	100	(12; 12)	10.7	100	100	202.8 (22.9)	10.9 (6.2)	47.7 (7.10)
	6	Control	B	22	100	(15; 7)	10.6	100	100	194.1 (20.3)	8.9 (6.9)	52.9 (11.0)
	7	CG	M	23	78	(7; 16)	10.5	72	78	202.0 (19.5)	15.9 (7.9)	53.8 (12.6)
	8	DI	M	23	83	(12; 11)	10.6	84	79	215.7 (25.3)	21.4 (10.0)	60.5 (10.1)
	9	Control	M	23	83	(11; 12)	10.7	63	58	210.4 (20.0)	15.4 (11.0)	58.1 (13.0)
2	1	CG	M	27	85	(12; 15)	10.5	78	70	218.7 (17.7)	18.9 (8.2)	57.3 (12.0)
	2	DI	M	27	44	(12; 15)	10.6	83	67	205.7 (19.6)	14.5 (8.5)	62.3 (13.7)
	3	Control	M	26	85	(18; 8)	10.4	82	68	209.8 (22.3)	15.6 (8.4)	51.0 (8.1)
	4	CG	M	27	56	(10; 17)	10.5	87	80	202.7 (20.9)	12.8 (7.9)	57.4 (13.7)
	5	DI	M	24	63	(11; 13)	10.5	80	87	207.6 (16.0)	10.1 (8.6)	57.3 (13.0)
	6	Control	M	25	100	(16; 9)	10.4	92	100	188.8 (16.1)	6.1 (6.8)	49.2 (13.1)
	7	CG	M	22	91	(8; 14)	10.4	90	85	213.3 (18.1)	16.7 (9.4)	58.6 (11.1)
	8	DI	M	23	61	(6; 17)	10.5	86	93	209.0 (17.9)	16.2 (6.2)	55.3 (14.3)
	9	Control	M	22	86	(11; 11)	10.7	68	68	212.2 (27.4)	16.3 (10.2)	54.3 (13.2)

Notes.[1]Condition: CG refers to the collaborative group work condition, DI refers to the direct instruction condition, and control refers to the control condition

[2]Program: M stands for mainstream classroom and B stands for sheltered bilingual classroom

[3] Among children identified as Hispanic

[4] Fourth-grade Illinois Standard Achievement Test reading scale score

[5] Gates-MacGinitie reading test score (corrected for guessing)

[6] Average number of objects that students correctly named per minute in rapid automatized naming task

Table 2

Factor loadings of SALT measures on five language factors

	Length	Syntactic Complexity	Fluency	Mazes	Errors & Omissions
Number of utterances	**.917**	-.143	-.011	.285	.142
Number of complete words	**.843**	.199	.020	.455	.099
Number of different words	**.909**	.250	.043	.183	.072
MLU in words	.082	**.956**	.037	.101	-.095
MLU in morphemes	.095	**.960**	.021	.072	-.111
Subordination index	.044	**.874**	.050	-.063	.086
Number of mazes	.220	.049	-.057	**.939**	.141
Utterance with mazes	.309	.022	-.020	**.896**	.130
Number of maze words	.214	.032	.006	**.939**	.109
Words per minute	.282	.142	**.750**	.056	-.220
Pauses	-.177	-.026	**.868**	-.095	.081
Errors	.019	-.062	-.052	.245	**.798**
Omissions	.187	-.022	-.035	.052	**.849**

Table 3

Attempts and outcomes in the frog story and an example of multi-link reasoning

	Attempts	Outcomes	An Example of Multi-link Reasoning Structure
	Initiating event: When the boy woke up, he found the frog was missing.		
1	The boy searched in the bedroom.	The boy did not find the frog.	C They went outside and looked. C *Nothing happened.*
2	The boy and the dog called for the frog outside the window.	The boy did not find the frog.	C They went to the trees. C *And nothing happened*. C They went to a beaver hole.
3	The boy and the dog searched outside in the forest.	The boy did not find the frog.	C (Th*) *She wasn't there*. C They went to look on the (bee) bee hole. C *And there was nothing there*.
4	The boy looked for the frog in a hole in the ground.	The boy did not find the frog.	C They look*ed everywhere, even on trees, even where the owl lives.
5	The boy searched for the frog in a hole in the tree.	The boy did not find the frog.	C (They lo*) they look*ed under rocks. C they looked over rocks. C they called him X his name.
6	The boy climbed up a rock to try to see the frog.	The boy did not find the frog.	C they even looked (on) on (UM) deer, *but the deer didn't want them to check on there*. C They looked on the water.
7	The boy and the dog climbed on a log to listen for the frog.	The boy found the frog and his family!!	C *And nothing was there*. C They look*ed over the log. C *Something was there*. C (Th*) finally they found the frog.

Note. In the example of multi-link reasoning structure, C refers to a C-unit generated by a child.

Table 4

Performance on pretests of reading comprehension and oral English proficiency

	Condition		
	CG (N=70)	DI (N=68)	Control (N=72)
Pretest	M (SD)	M (SD)	M (SD)
Reading comprehension	16.24 (8.45)	15.38 (8.92)	13.46 (9.32)
Oral English proficiency	53.71 (12.03)	54.86 (12.20)	52.32 (11.32)
Fourth-grade ISAT reading	211.34 (21.00)	209.33 (21.77)	203.50 (20.15)

Note. CG refers to the collaborative group work condition and DI refers to the direct instruction condition.

Table 5

Descriptive statistics of language production and story elements in each intervention condition

	Condition[1]		
	CG (N= 68) M (SD)	DI (N= 67) M (SD)	Control (N= 70) M (SD)
Language production[2]			
Length	49.7 (20.3)	48.8 (23.4)	51.6 (16.6)
Syntactic complexity	**54.6** (20.7)	50.5 (18.7)	46.2 (19.6)
Fluency	49.0 (17.9)	50.7 (21.7)	51.2 (16.9)
Mazes	**52.0** (23.3)	49.6 (18.5)	48.3 (18.1)
Errors & Omissions	47.8 (17.1)	47.3 (18.6)	**54.9** (23.4)
Story elements[3]			
Setting	2.57 (0.94)	2.52 (0.93)	2.39 (0.77)
Initiating events	2.50 (0.70)	2.13 (0.83)	2.41 (0.81)
Internal responses	5.78 (3.52)	5.53 (3.87)	5.08 (3.09)
Attempts	3.97 (2.43)	3.69 (2.34)	3.43 (2.06)
Actions	5.24 (2.70)	4.54 (2.67)	6.09 (2.52)
Outcomes	3.49 (2.11)	2.57 (1.77)	2.90 (1.49)
End State	3.12 (0.94)	3.22 (0.95)	2.86 (0.92)
Minor events	11.81 (6.27)	12.15 (6.95)	12.59 (4.97)
Essential Story Elements[4]	**21.44** (6.68)	19.67 (7.70)	19.07 (5.87)

Note. [1] CG refers to the collaborative group work condition and DI refers to the direct instruction condition.

[2] Means and standard deviations of percentaged language factor scores.

[3] Means and standard deviations of number of non-repetitive C-units.

[4] Essential Story Elements include setting, initiating event, internal response, attempt, outcome, and end state.

Table 6

Multinomial logistic regression of length of multi-link reasoning chains

	Odds Ratio (95% CI)	
Chain length	Collaborative Groups vs. Control	Direct Instruction vs. Control
1-2 links	1.30 (0.56 to 3.00)	2.67 (1.12 to 6.40)
3-4 links	1.03 (0.36 to 2.91)	0.82 (0.25 to 2.66)
5-7 links	15.18 (1.53 to 150.85)	6.36 (0.54 to 75.03)

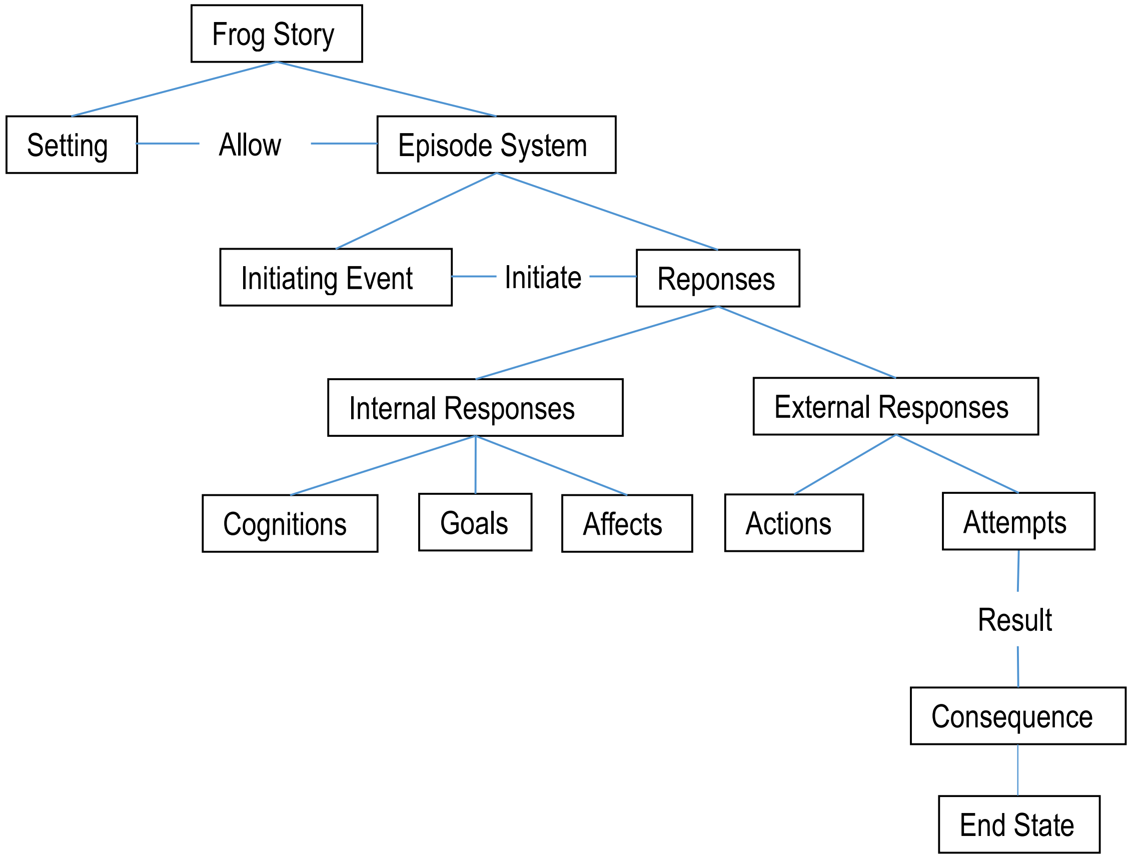
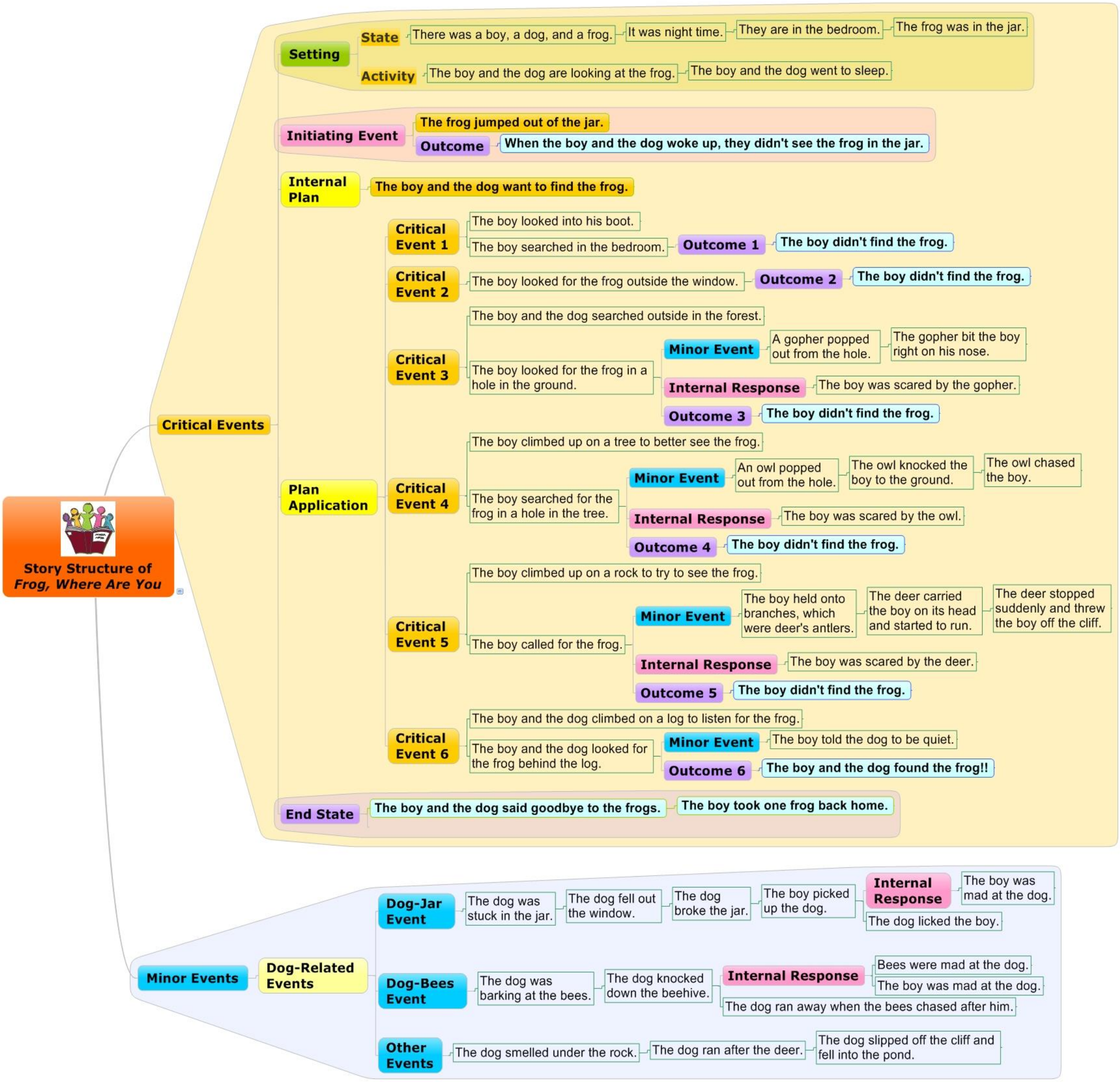


Figure 1. Story gammar of the *Frog, Where Are You* story



Story Structure of *Frog, Where Are You*

Critical Events

Setting

State

There was a boy, a dog, and a frog.

It was night time.

They are in the bedroom.

The frog was in the jar.

Activity

The boy and the dog are looking at the frog.

The boy and the dog went to sleep.

Initiating Event

The frog jumped out of the jar.

Outcome

When the boy and the dog woke up, they didn't see the frog in the jar.

Internal Plan

The boy and the dog want to find the frog.

Plan Application

Critical Event 1

The boy looked into his boot.

The boy searched in the bedroom.

Outcome 1

The boy didn't find the frog.

Critical Event 2

The boy looked for the frog outside the window.

Outcome 2

The boy didn't find the frog.

Critical Event 3

The boy and the dog searched outside in the forest.

The boy looked for the frog in a hole in the ground.

Minor Event

A gopher popped out from the hole.

The gopher bit the boy right on his nose.

Internal Response

The boy was scared by the gopher.

Outcome 3

The boy didn't find the frog.

Critical Event 4

The boy climbed up on a tree to better see the frog.

The boy searched for the frog in a hole in the tree.

Minor Event

An owl popped out from the hole.

The owl knocked the boy to the ground.

The owl chased the boy.

Internal Response

The boy was scared by the owl.

Outcome 4

The boy didn't find the frog.

Critical Event 5

The boy climbed up on a rock to try to see the frog.

The boy called for the frog.

Minor Event

The boy held onto branches, which were deer's antlers.

The deer carried the boy on its head and started to run.

The deer stopped suddenly and threw the boy off the cliff.

Internal Response

The boy was scared by the deer.

Outcome 5

The boy didn't find the frog.

Critical Event 6

The boy and the dog climbed on a log to listen for the frog.

The boy and the dog looked for the frog behind the log.

Minor Event

The boy told the dog to be quiet.

Outcome 6

The boy and the dog found the frog!!

End State

The boy and the dog said goodbye to the frogs.

The boy took one frog back home.

Minor Events

Dog-Related Events

Dog-Jar Event

The dog was stuck in the jar.

The dog fell out the window.

The dog broke the jar.

The boy picked up the dog.

Internal Response

The boy was mad at the dog.

The dog licked the boy.

Dog-Bees Event

The dog was barking at the bees.

The dog knocked down the beehive.

Internal Response

Bees were mad at the dog.

The boy was mad at the dog.

The dog ran away when the bees chased after him.

Other Events

The dog smelled under the rock.

The dog ran after the deer.

The dog slipped off the cliff and fell into the pond.

Figure 2. Coding scheme for the *Frog, Where Are You* story

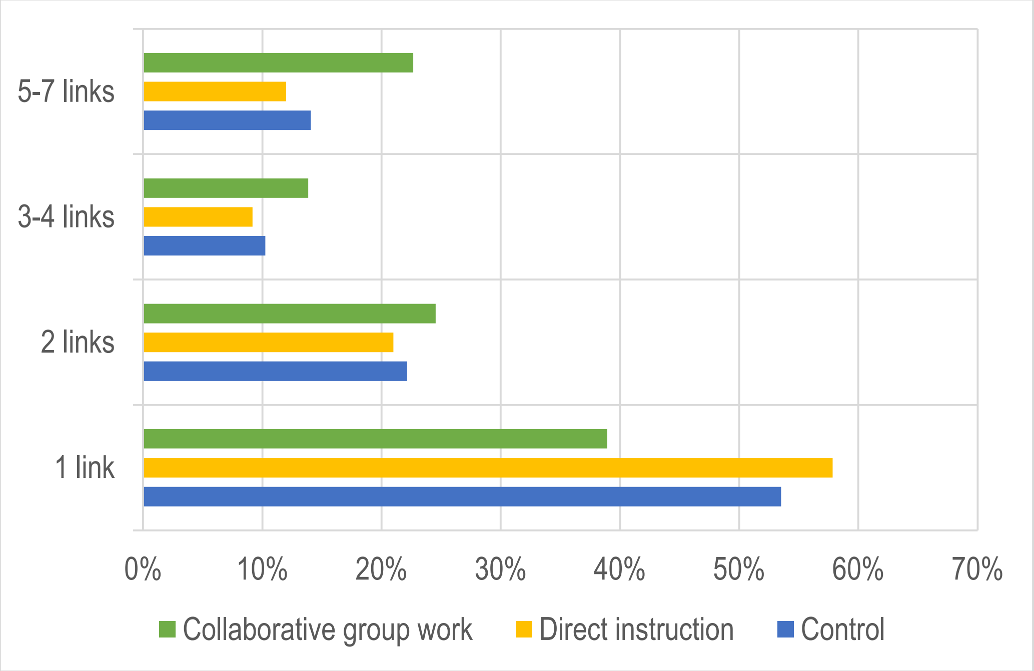


Figure 3. Probability of generating multi-link reasoning chains of different lengths by intervention condition

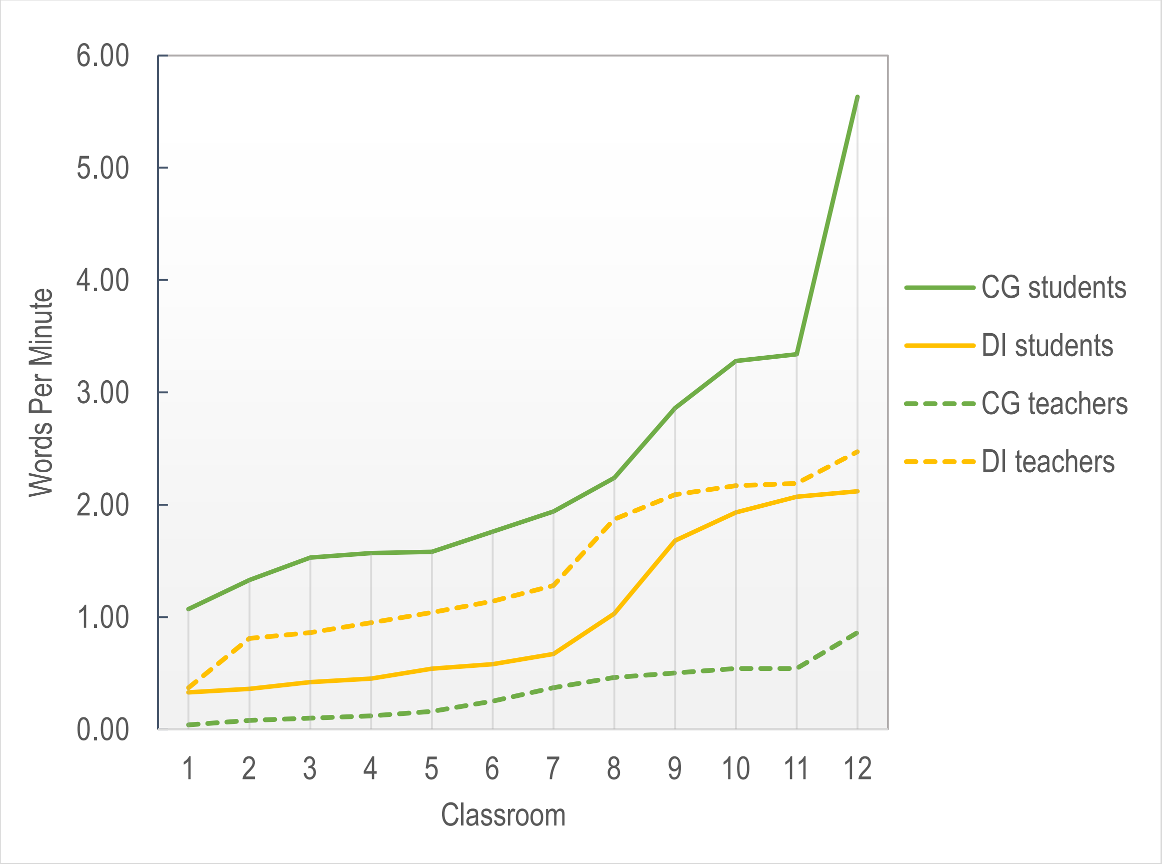


Figure 4. Students' and teachers' academic vocabulary words per minute by classroom within condition