Discourse Processes

ISSN: 0163-853X (Print) 1532-6950 (Online) Journal homepage: http://www.tandfonline.com/loi/hdsp20

Children's Productive Use of Academic Vocabulary

Shufeng Ma, Jie Zhang, Richard C. Anderson, Joshua Morris, Kim Thi Nguyen-Jahiel, Brian Miller, May Jadallah, Jingjing Sun, Tzu-Jung Lin, Theresa Scott, Yu-Li Hsu, Xin Zhang, Beata Latawiec & Kay Grabow

To cite this article: Shufeng Ma, Jie Zhang, Richard C. Anderson, Joshua Morris, Kim Thi Nguyen-Jahiel, Brian Miller, May Jadallah, Jingjing Sun, Tzu-Jung Lin, Theresa Scott, Yu-Li Hsu, Xin Zhang, Beata Latawiec & Kay Grabow (2017) Children's Productive Use of Academic Vocabulary, Discourse Processes, 54:1, 40-61, DOI: 10.1080/0163853X.2016.1166889

To link to this article: http://dx.doi.org/10.1080/0163853X.2016.1166889

Accepted author version posted online: 04 Apr 2016.
Published online: 04 Apr 2016.

DISCOURSE PROCESSES
2017, VOL. 54, NO. 1, 40–61
http://dx.doi.org/10.1080/0163853X.2016.1166889

Routledge
Taylor & Francis Group

Children's Productive Use of Academic Vocabulary

Shufeng Ma[a], Jie Zhang[b], Richard C. Anderson[a], Joshua Morris[a], Kim Thi Nguyen-Jahiel[c], Brian Miller[d], May Jadallah[e], Jingjing Sun[f], Tzu-Jung Lin[g], Theresa Scott[h], Yu-Li Hsu[a], Xin Zhang[a], Beata Latawiec[i], and Kay Grabow[j]

[a]Center for the Study of Reading University of Illinois at Urbana-Champaign; [b]Department of Educational Administration, Leadership, and Research Western Kentucky University; [c]College of Engineering University of Illinois at Urbana-Champaign; [d]Department of Elementary Education Towson University; [e]School of Teaching and Learning Illinois State University; [f]Department of Teaching and Learning University of Montana; [g]Department of Educational Studies The Ohio State University; [h]Booker T. Washington STEM Academy Washington Elementary School, Champaign, Illinois; [i]Department of Counseling, Educational Leadership, Educational and School Psychology Wichita State University; [j]Thomas Paine Elementary School, Urbana, Illinois

ABSTRACT

Instructional influences on productive use of academic vocabulary were investigated among 460 mostly African American and Latina/o fifth graders from 36 classrooms in eight public schools serving low-income families. Students received a 6-week unit on wolf management involving collaborative group work (CG) or direct instruction (DI). The big question that students tried to answer during the unit was whether a community should be permitted to destroy a pack of wolves. In an individual oral interview about an analogue to the wolf question, whether whaling should be allowed, both CG and DI students used more general and domain-specific academic vocabulary from the Wolf Unit than uninstructed control students. CG students used more general academic vocabulary in the whale interview than DI students, and this was mediated by the CG students' greater use of general academic vocabulary in classroom dialogue during the Wolf Unit. These results suggest that CG is an effective instructional approach to promote acquisition and productive use of academic vocabulary for children from underserved communities.

Introduction

Mastery of academic vocabulary is necessary for comprehension of content area textbooks, technical material in trades and professions, and newspaper articles on current events and is no doubt important for thinking, speaking, and writing about all manner of topics. In research on growth in knowledge of academic vocabulary, the test for whether or not students know academic words has almost always involved recognizing the words or selecting definitions of the words from among several choices. Almost no research has been done on *productive use* of academic vocabulary words (Pearson, Hiebert, & Kamil, 2007).

In a productive use of a vocabulary item, an individual is able to say or write the word in contrast to merely being able to recognize it. As we use the term, *productive use* has the further requirement of a spontaneous or unprompted use. Thus, although a task in which children rapidly name common objects meets one requirement of productive use of words, the pictures of the objects, and the instructions to name them, strongly prompt the words. The meaning of the term *productive use* is satisfied when a child (1) articulates or writes a word and (2) does so spontaneously, as when telling a

CONTACT Shufeng Ma, MA sma10@illinois.edu Center for the Study of Reading, University of Illinois at Urbana-Champaign, 157 Children's Research Center, 51 Gerty Drive, Champaign, IL 61820.

story or writing an essay. A test for productive use of a word need not be "natural" in the sense, for instance, of overhearing a child use the word on the playground. The use can be elicited as long as the child has plenty of degrees of freedom in how to respond. Ideally, the elicitation would afford but not require use of the word. A robust assessment of productive use will evaluate whether the child can use the word in a context different from the one in which it was acquired. The term *productive use* is comparable with Laufer and Nation's (1999) term *free productive use* that they contrasted with *controlled productive use*. Supplying a suitable word in the following context is an example of controlled productive use: "The garden was full of fra_____ flowers."

The present study examined instructional influences on productive use of academic vocabulary. The influence of two methods, direct instruction (DI) and collaborative group work (CG), was compared with African American and Latina/o children, the two largest groups of underserved children in the United States. The National Assessment of Educational Progress reports that these students continue to lag in reading and writing (Salahu-Din, Persky, & Miller, 2008). Students from low income, minority homes often lack familiarity with lexical, grammatical, and discourse features of an academic voice (Scarcella, 2003; Snow & Uccelli, 2009). Among underserved children who are second language learners, there is a large gap between receptive language (listening and reading) and expressive language (speaking and writing), with expressive language lagging as much as 1 standard deviation behind receptive language, according to one recent study (Keller, Troesch, & Grob, 2015). Hence, there is a pressing need to identify instructional methods for underserved children that expand knowledge of academic vocabulary words and create functional contexts for productive use of the words.

Challenge of academic vocabulary

Academic vocabulary is an important component of academic language, a register of English used in schools and universities that is critical for academic success (Corson, 1997; Scarcella, 2003; Snow, 2010). Academic vocabulary can be classified into two categories: domain-specific vocabulary, also called technical vocabulary (Hiebert & Lubliner, 2008) or Tier 3 words (Beck, McKeown, & Kucan, 2002), and general academic vocabulary (GAV; e.g., Bailey, 2006; Hiebert & Lubliner, 2008). Domain-specific vocabulary refers to technical words that are necessary for understanding and expressing key concepts within a domain, such as *mean* and *standard deviation* within the domain of statistics. GAV refers to broadly useful words that appear in the discourse of many disciplines, such as *affect*, *decline,* and *provide*.

Knowing domain-specific vocabulary is indispensable for proficiency in science and other technical subjects. Domain-specific words pose a challenge for students because these words tend to be labels for unfamiliar concepts (Bravo & Cervetti, 2008). Although domain-specific vocabulary is challenging for all students, GAV also presents a challenge for English language learners and students from low-income and less-educated families. These students depend on schools for exposure to academic words, which are "usually non-concrete, low in imagery, low in frequency, and semantically opaque" (Corson, 1997, p. 696) and therefore difficult to learn.

Incidental learning of words while reading plays an essential role in vocabulary growth (Nagy & Anderson, 1984; Stanovich, 1986). However, research suggests that level of word abstractness influences vocabulary learning (Schwanenflugel et al., 1997), which makes acquisition of academic vocabulary more demanding than acquisition of other words. There is little evidence of incidental learning of complex and abstract words from one or two exposures during normal reading (Nagy, Anderson, & Herman, 1987). Because academic vocabulary is often complex and abstract, studies suggest that incidental learning is not enough for students to acquire new academic words among second language learners and other students with limited exposure to academic discourse (Carlo et al., 2004). Many have argued that to accelerate these students' growth in literacy and content knowledge, academic vocabulary should be taught explicitly (e.g., Snow, 2010).

In recent years there has been an increase in research evaluating instructional interventions designed to teach academic vocabulary to students from linguistically diverse backgrounds. Some interventions have focused on explicit teaching of sets of generally useful academic vocabulary words (e.g., Beck, McKeown, & Kucan, 2002; Carlo et al., 2004; Snow, Lawrence, & White, 2009; Lesaux, Kieffer, Faller & Kelley, 2010; Townsend & Collins, 2009), whereas others have integrated vocabulary teaching into science or literacy instruction (e.g., August, Branum-Martin, Cardenas-Hagan, & Francis, 2009; Vaughn et al., 2009). These interventions show some promising effects on English vocabulary knowledge and reading comprehension (although see Marulis & Neuman, 2010) and on content knowledge, regardless of the student's first language.

Academic vocabulary acquisition through classroom discussion

Research suggests that vocabulary learning is most likely to happen when students have multiple exposures to vocabulary words in varied and meaningful contexts (e.g., Kelley, Lesaux, Kieffer, & Faller, 2010). An interactive learning format, therefore, may increase students' chances of encountering and using academic vocabulary and provide more opportunities for them to negotiate meanings in different contexts (Oxford, 1997). In a review of effective approaches to language instruction, Ellis (2005) pointed out that, as compared with DI, small group work increases classroom talk and involves students in a greater variety of speech acts. Mol, Bus, and de Jong (2009) found that an interactive storybook reading program resulted in an 8% increase in children's expressive vocabulary. Carlisle, Fleming, and Gudbrandsen (2000) reported that involvement in discussion and hands-on activities in fourth and eighth grade science led to vocabulary growth, particularly when students already had partial knowledge of the words.

Classroom discussion may facilitate the acquisition of academic vocabulary because discussion often creates a high level of involvement, perhaps leading to deep processing of information. Discussion requires understanding words in context, supplementing the understanding that can be gleaned from definitions. Barron and Melnick (1973) compared three approaches to teaching a set of biology concepts to 10th graders: student-led small group discussion, teacher-led whole-class discussion, and individual completion of worksheets. Results showed that both types of discussion led to higher scores on assessments of the biology vocabulary than did individual exercises, but there was no difference between the two discussion types. Stahl and Vancil (1986) found similar results with fifth grade students who were learning a set of meteorology concepts.

These findings suggest that discussion can play a role in the acquisition of domain-specific academic vocabulary, at least when the words have been explicitly taught and integrated into discussion. Some form of discussion is often one component of interventions to improve vocabulary. For example, Lesaux et al. (2010) implemented small-group work in addition to whole-class and independent activities to teach sixth-grade students a list of academic words, and Snow et al. (2009) incorporated argumentative discourse as a complement to explicit vocabulary instruction. The role small group work and argumentative discourse play in academic vocabulary acquisition has yet to be clearly established, however, because in previous studies these practices have been intermingled with other types of vocabulary instruction.

Social factors in classroom learning

Social dynamics are important to classroom learning outcomes. According to a comprehensive review of the literature by Howe and Mercer (2007), children's social histories and characteristics such as popularity, status in the classroom, temperament, and social experience at home and in the community impact the quality of interaction in the classroom. Murphy and Faulkner (2000) reported that popular children were more likely to maintain successful collaboration as compared with unpopular children, because popular children were more capable of following rules (e.g., turn-taking) and more strategic in verbal and nonverbal communication, such as providing elaborated arguments and monitoring group members' facial expressions. Lin et al. (2015) found that students' status in the classroom social network mediated the effectiveness of collaborative discussions in improving relational thinking. Students

centered within the network (i.e., students with high status) provided both more support and more challenges to their peers and played a role in creating a harmonious and productive experience for all.

Students' acquisition of academic vocabulary is likely to depend on the quality of interaction. Talkativeness, leadership qualities, and social status or, conversely, social insecurity likely influence students' willingness to try new words. Toward a more complete understanding of children's vocabulary development, the present study considers both social and cognitive factors that affect peer interaction. To our knowledge, this study is the first that seeks to understand the influence of social factors on children's acquisition of academic vocabulary.

Rationale for the present study

The present study investigated how CG and whole-class DI impact underserved children's productive use of academic vocabulary, as compared with a control condition in which students continued regular instruction. As a vehicle for investigating types of instruction, we developed a curriculum unit, called the Wolf Reintroduction and Management Unit, intended to be conceptually rich and intellectually stimulating that integrates language arts, science, math, and social studies (Jadallah et al., 2009). Students played the role of officials at the Wolf Management Agency who are responsible for deciding whether or not a community should be allowed to eradicate a wolf pack spotted nearby. Three domains of knowledge were introduced to discuss this sensitive policy issue taking into account potential effects on the local ecosystem, the town's economy, and public policy. Key terms were defined as they appeared in the text and were repeated in the margins and in the glossary at the end of each booklet. In an individual oral transfer task that took the form of an interview, students heard a statement about the pros and cons of whaling and then were asked for their own position on whether or not whaling should be allowed. The transcripts of students' interviews were searched for both general and domain-specific academic vocabulary introduced in the Wolf Unit.

This study was motivated by several gaps in current research. First, to our knowledge, the present study is among the few to investigate the productive use of academic vocabulary. Second, despite the generally recognized importance of multiple and varied exposure to words, few studies have addressed how conditions of teaching and learning academic vocabulary words affect the likelihood of students using the words in contexts different from the one in which they were learned. Third, although the value of discussion in promoting academic vocabulary is widely appreciated, little research has been dedicated to determining how and why it works. Finally, except for facets of language ability, research has yet to explore how other individual and social characteristics of students influence their uptake and use of academic vocabulary.

To fill these gaps, this study addresses three research questions. First, do collaborative small groups and whole-class DI have differential effects on student productive use of academic vocabulary? Our expectation was that CG would exceed DI and the control that continued regular instruction. One basis for this expectation is that CG provides more numerous and varied opportunities to use academic vocabulary words. Second, does classroom dialogue affect students' productive use of academic vocabulary? We expected that frequency of academic vocabulary during classroom talk would mediate students' productive use of academic vocabulary in the later transfer task. Third, do students' social characteristics influence their use of academic vocabulary words? We anticipated that socially centered students (and talkative students judged by their peers to be leaders or to have good ideas) would use more academic vocabulary than socially peripheral students because centered, or high status, students are more active in classroom activities and may be expected take the initiative and try new words.

Methods

Participants

The sample consisted of 460 fifth-grade students enrolled in 36 classrooms in eight public schools in low-income school districts in central and northern Illinois, who participated in this project across two

academic years. Each year 18 classrooms were recruited, 9 classrooms with a predominant enrollment of African American students and another 9 classrooms with a predominant enrollment of Spanish-speaking English language learners. Classrooms within triples of classrooms matched on demographic characteristics and previous academic performance were randomly assigned to one of three intervention conditions: CG, DI, or wait-listed control that continued regular instruction and received the intervention in the following semester after the data were collected.

The 460 participants included 160 CG students, 153 DI students, and 147 control students, mainly African American (41%) and Latina/o (49%), approximately balanced by gender (245 girls and 215 boys), average age 10.7 ($SD = .5$). Depending on the school, from 79% to 99% of the students were registered for free or reduced lunch. The first language of most of the Latinas and Latinos was Spanish and, according to a home survey 84% of them spoke Spanish or a mixture of Spanish and English with their parents. Thirty-two students (7%) had an Individualized Education Program and received special services, distributed among conditions as follows: 13 CG students, 12 DI students, and 7 control students.

Procedure

Students in the CG and DI conditions studied a 6-week-long Wolf Reintroduction and Management Unit that addressed a socioscientific policy issue faced by an imaginary community that had requested permission to hire hunters to kill wolves that alarmed many of its citizens (Jadallah et al., 2009). The unit was divided into three sections, each incorporating an important perspective on the complicated issue of wolves, to cultivate students' ability to discern different aspects of problems and understand interrelationships and trade-offs. The three sections were ecosystem, economy, and public policy. Although killing the wolves may be favored by most residents in the community (public policy), doing so would alter the food web (ecosystem), which would impact community businesses (economy). Each section of the unit was explained in an information booklet and expanded in an activity booklet. Information booklets provided students with essential concepts. Unlike most readings for middle grade students, the booklets had an argument structure that contrasted opposing viewpoints.

CG and DI teachers attended parallel 2-day workshops to learn about their assigned instructional approach to the Wolf Unit. CG teachers learned the theory and research base for CG, how to facilitate Collaborative Reasoning discussions, and best practices for CG (Gillies, 2007). DI teachers learned the theory and research for whole-class teacher directed methods and the best practices for DI (Institute of Educational Sciences, 2007). Teachers saw videos of the Wolf Unit being taught by the method they were supposed to use.

The intervention encompassed about 22 class sessions over 6 weeks. In the CG condition the Wolf Unit was implemented in a modified jigsaw format. Students were divided into three heterogeneous "expert" groups to study one of the three domains: ecosystem, economy, or public policy. At the beginning of the unit, each expert group in the CG condition had a Collaborative Reasoning discussion in which they talked about their initial opinions on the central question: whether the community should be permitted to hire professional hunters to kill the wolves. Students were encouraged to take positions and defend their positions with arguments and evidence to make the policy decision. Then each expert group studied its assigned domain through readings and group activities. After 4 weeks each expert group taught the rest of the class what they had learned by presenting posters. After the poster presentations, students were assigned to new groups that contained experts from each knowledge domain. The new groups had a second Collaborative Reasoning discussion to once again reflect on the central question.

In the DI condition the Wolf Unit involved whole-class teacher-directed instruction and individual seatwork but no small group discussions or other CG or poster presentations. Students learned all three subtopic domains through teacher explanation and teacher-managed whole-class discussion. Students read the information booklets and completed the activity booklets individually at their seats.

The CG version of the Wolf Unit printed materials contained five sections: introductory packet, wolves in the United States, information booklet on one subtopic domain, activity booklet on that subtopic, and poster booklet on the same subtopic. Each group within a CG classroom had a different set of printed materials focusing on one of the three subtopics. The printed materials for the DI version of the Wolf Unit contained eight sections: introductory packet, wolves in the United States, and three pairs of information and activity booklets on the three subtopics. The students in DI classrooms all had the same comprehensive set of printed materials. The information booklets for the two conditions were identical except for the subheadings. The subheadings in the DI version were written in declarative form in keeping with the authoritative stance of DI, whereas the subheadings in the CG version were written in interrogative form in keeping with the inquiry stance of CG.

DI students were exposed to all three subtopics, whereas CG students studied one subtopic. The study was set up in this way to keep time studying the Wolf Unit approximately equal for the two intervention conditions, because DI is more efficient in covering the curriculum than CG. Moreover, although this instructional design resulted in unbalanced exposure to domain-specific vocabulary, it served the goal of a thoroughgoing implementation of contrasting instructional approaches. The jigsaw design enabled CG students to have distinctive knowledge to contribute to discussions, which presumably helped groups more readily appreciate the value of achieving shared understanding and constructive challenging of one another. CG students learned concepts in one domain of knowledge and later served as "teachers" when the group talked about their domain but also played the role of learners when talk turned to the domains less familiar to them.

In the wait-listed control condition, classrooms did not study the wolf curriculum during the intervention period but continued regular instruction. Control classrooms had the opportunity to experience the Wolf Unit in the following semester after the data were collected.

Initial language measures. Pretest assessments included the Gates-MacGinitie reading comprehension test (MacGinitie, MacGinitie, Maria, & Dreyer, 2000) and a speeded object naming task (Snodgrass & Vanderwart, 1980) intended to assess children's basic oral English proficiency. Reading scores were corrected for guessing (right – wrong/3), which improved reliability and predictive validity. Oral English proficiency was indicated by number of common objects, such as bike, car, and bus, correctly named per minute.

Initial measures of social characteristics. A sociometric questionnaire elicited peer nominations and peer ratings to provide assessments of popularity, friendships, social status, talkativeness, quietness, leadership, and reputation for having good ideas. Children's talkativeness was calculated by deducting the number of classmates' nominations for quietness from the number of nominations for having a lot to say in class discussions. Three indices of students' social status, or *centrality*, were derived from friendship nominations using social network analysis (Butts, 2008). *Indegree centrality* represents an individual's popularity; it refers to how often a student was nominated by his or her classmates as a friend. *Betweenness centrality* represents how often a student was nominated as the common friend of two other unconnected students. *Information centrality* represents how far away a student is from every other student in the friendship network. Peer-liking ratings were summed ratings on a five-point Likert scale of how much children liked to play with each of the other children in the class. To adjust for differences in class size, all social measures were divided by the number of children in the class.

Whale policy transfer interview. In the individual oral transfer task that followed the Wolf Unit, each student heard a 386-word statement about the pros and cons of whaling and then was asked to explain his or her own position on whether whaling should be allowed. The examiner gave standardized prompts if the student stopped short of providing a complete argument (for more detail see Appendix A). Student responses to the whale question were audio recorded and transcribed verbatim following the Systematic Analysis of Language Transcripts conventions (Miller & Chapman, 2010).

Because of time and resource constraints, only about 60% of the students were interviewed to present their opinions on the whaling question. Interviewed first were "target students," so-called

because these were the students on whom the video camera was trained throughout the unit, selected from each class with the help of teacher to be a representative sample of the class in terms of gender, ethnicity, academic performance, and talkativeness. In CG classrooms target students studied the ecosystem knowledge domain and who worked together every day as a group. In DI classrooms target students were a representative sample of the class, selected according to the same criteria as CG target students, who sat together and were videotaped every day but who did not meet for group activities. In control classrooms, there were nominal target students, also selected at the beginning of the study to be a representative sample of the class in terms of gender, ethnicity, academic performance, and talkativeness but who were not videotaped during the period of the intervention and did not work together as a group.

In the week after the intervention was completed, students were pulled out of the class one by one to be interviewed. The interview was completed one class at a time. Research assistants who conducted the interview were given a list of randomly ordered names of all the students in the class, with target students on the top and nontarget students filling out the rest of the list. Target students were interviewed first and then as many nontarget students as possible following the order of the list. Eventually, in CG classrooms we interviewed 88 ecosystem students, 36 economy students, and 36 public policy students. In DI classrooms 153 target and nontarget students were interviewed, whereas in control classrooms 147 target and nontarget students were interviewed.

We created a two-level mixed-effects regression model to compare the pretest performance and social characteristics of students who were interviewed about the whale question and students who were not. Pretest reading comprehension, object naming, talkativeness, good idea nomination, leader nomination, and peer liking rating were dependent variables, and intervention condition and whether or not the student received the whale interview were fixed effects. Classroom was entered as a random effect to account for variance due to teacher or cohort. There was no difference in the object naming of students who were interviewed and students who were not interviewed, $F(1, 725) = .35$, $p = .56$. Students who took the interview had higher pretest reading comprehension scores than those who did not, $F(1, 725) = 9.27$, $p < .01$. However, this difference was observed in all three conditions, as indicated by the nonsignificant condition effect, $F(2, 725) = .84$, $p = .43$. Students who took the interview were also more talkative, $F(1, 725) = 6.63$, $p = .010$, more likely to be nominated as having good ideas, $F(1, 725) = 10.99$, $p < .01$, and more likely to be nominated as leaders, $F(1, 725) = 12.35$, $p < .01$; however, these differences applied to all three conditions. Students who did and did not take the interview were equally liked by their classmates, $F(1, 725) = .10$, $p = .75$. Because of time constraints, research assistants were instructed to skip students on the randomly ordered name list who were absent. Students whose name was called but were not interviewed either were absent from school on that day or were receiving special instruction somewhere else in the school as part of their Individualized Education Program. Apparently, children with high reading scores and children considered to be talkative or to have good ideas or leadership qualities were less likely to be absent from class.

Identifying academic vocabulary in the wolf unit printed materials

Following Bailey's (2006) classification of academic vocabulary, student use of general and domain-specific academic vocabulary was investigated in this study. A list of 60 domain-specific words that convey the core concepts of the Wolf Unit were identified by the authors of the curriculum. There were 25 Ecosystem words (e.g., *food web, predator, species*), 17 Economy words (*tourism, economy, ranching*), and 18 Public Policy words (*advocate, common good, representation*). To determine the print exposure of these domain-specific words for students in the CG expert groups and the DI students, the Wolf Unit materials for each group were searched to determine the frequency of occurrence each of the words (for more detail see Appendix B). The inflectional and derivational variants of words were included in the counts. For example, the exposure frequency of the word *reintroduce* would include that of *reintroducing, reintroduced,* and *reintroduction*. Expert groups in the CG classrooms had a high rate

of exposure to the domain-specific words in their own domain but less exposure to the words outside this domain. Four domain-specific words from the Wolf Unit appeared in the 386-word statement that the examiner read to the child in the whale transfer task, namely *opinion*, *population*, *endangered*, and *tradition*.

The identification of GAV in the Wolf Unit began with two lists of academic words. Coxhead's (2000) Academic Word List contains 570 word families that university students frequently encounter in textbooks, excluding any of the first 2,000 most frequent English words. Shoebottom's (2008) list of 1,040 general academic words was generated from a corpus of words that second language learners are expected to acquire by their second or third year of intensive English study. A search for the general vocabulary words from these two lists in the printed Wolf Unit materials netted 250 words that appeared at least once.

Coxhead's and Shoebottom's lists are skewed toward the vocabulary needed by college students and may not capture all GAV useful for fifth graders to know. To see if more words should be included, we examined all the low frequency words in the Wolf Unit materials and found 75 more words that we judged met the criteria for GAV. The first 50 most frequent GAV words are presented in Appendix C.

The printed materials assigned to be read during the Wolf Unit were somewhat different for DI students and for subgroups of CG students. These variations had to be taken into consideration in estimating the opportunity to learn academic vocabulary words. As shown in Appendix B, CG subgroups had a greater opportunity to learn the domain-specific words in their domains of expertise. A search of the different versions of the Wolf Unit printed materials showed that DI students, who were supposed to read all the printed materials in the unit except the poster booklets, had exposure to a greater range of GAV words ($N_{\text{CG\_Ecosystem}} = 286$; $N_{\text{CG\_Economy}} = 272$; $N_{\text{CG\_PublicPolicy}} = 306$; $N_{\text{DI}} = 325$).

Identifying academic vocabulary in whale interviews

The whale interview transcripts were searched for the 60 domain-specific words and the 325 general academic words using the text search query and word frequency function of NVivo qualitative data analysis software (QSR International Pty Ltd. Version 10, 2012). Of the 325 GAV words that appeared at least once in the Wolf Unit printed materials, 88 words were used by children in their oral whale interview responses. An additional 70 words from Coxhead's or Shoebottom's lists that do *not* appear in the Wolf Unit printed materials were also identified in the interviews. We used vocabulary *types* instead of *tokens* in the following analyses because types represent breadth of word knowledge. Because the frequency of vocabulary types is a count variable conforming to the Poisson distribution, mixed-effects Poisson regression models were constructed to examine condition differences in the use of general or domain-specific academic vocabulary in the whale transfer interviews.

Results

Pretest language and social measures

Separate two-level regression analyses were conducted to check whether there was a difference between the CG condition, DI condition, and control condition in pretest reading comprehension or pretest object naming, with gender and ethnicity and intervention condition as fixed effects. Classroom was entered as a random factor at the second level to account for teacher or cohort effects. The results indicated no significant condition difference in reading comprehension, $F(2, 420) = .76$, $p = .47$, or object naming, $F(2, 420) = .69$, $p = .50$. There was a marginal gender difference in reading comprehension; Girls had slightly higher reading scores than boys, $F(1, 420) = 3.05$, $p = .082$. Latina/o children were slower naming common objects than African American children, mean difference $= -8.84$, $t(420) = -5.60$, $p < .001$. No ethnic difference was found in reading comprehension, and no gender difference was found in object naming.

Two-level regression analyses were performed to evaluate differences in social characteristics with gender, ethnicity, and intervention condition as fixed effects and classroom as a random variable. Dependent variables were talkativeness, good idea nominations, leadership nominations, indegree centrality, betweenness centrality, information centrality, and peer-liking rating. No condition difference was found in any of these indices. Girls were found to be less talkative than boys, $F(1, 420) = 8.18$, $p < .01$. However, girls were more popular than boys, $F(1, 420) = 12.49$, $p < .001$, more likely to be nominated as leaders, $F(1, 420) = 27.25$, $p < .001$, more likely to be nominated as having good ideas, $F(1, 420) = 18.11$, $p < .01$, had higher social status as indicated by information centrality, $F(1, 420) = 17.44$, $p < .001$, and were more liked by classmates, $F(1, 420) = 32.00$, $p < .001$. African American children were considered by their peers to be more talkative than Latina/o children, $t(420) = 1.98$, $p = .048$.

Uses of academic vocabulary in response to whale question

A total of 158 general academic word types and 31 domain-specific word types were present in children's whale policy interviews. Eighty percent of the students used one or more types of general academic words for a total frequency of 1,537. Forty-four percent of the students used one or more types of domain-specific words for 1,009 total uses.

Among the 158 GAV words occurring in the whale interview, 17 words were used more than 20 times: *reason, agree, disagree, affect, instead, fair, spend, argument, increase, decrease, survive, since, attack, amount, environment, fault,* and *harm*. Domain-specific vocabulary occurring in the whale interview included 16 ecosystem words, 5 economy words, and 10 public policy words. Twelve words were used more than 10 times in the interviews, including 7 ecosystem words (*ecosystem, endangered, extinct, food web, nature, population, species*), two economy words *(economy, tourism*), and three public policy words (*majority, opinion, tradition*). Use of ecosystem words was observed among 64% of CG ecosystem students, 65% of CG economy students, 79% of CG public policy students, 50% of DI students, and 40% of control students. It would seem, therefore, that among CG students more use of ecosystem words than economy and public policy words was not due to more ecosystem students being interviewed.

Table 1 presents means and standard deviations of the occurrence of the two types of academic vocabulary words by intervention condition. Students in CGs showed a higher frequency of general academic words and a slightly higher frequency of domain-specific words than students who received DI. Both CG and DI students used more academic words of both types than control students. Ecosystem words (65%) were most frequently used, economy words (3%) were least used, and the use of public policy words (35%) was in the middle.

Table 2 presents the zero-order Pearson correlations between use of academic vocabulary in the whale interview and pretest language and social measures. Both reading comprehension and object naming were significantly correlated with the use of general academic words and domain-specific words, as were several of the social measures.

Factors predicting use of GAV. Mixed-effects Poisson regression models were constructed for the occurrence of GAV words in the whale interview. Fixed effects were gender, ethnicity, and intervention

Table 1. Means and Standard Deviations of Use of Academic Vocabulary Types in Whale Interview by Intervention Condition.

		Condition		
Variable		CGs ($n = 160$)	DI ($n = 153$)	Control Condition ($n = 147$)
Total occurrences of academic vocabulary		4.79 (3.51)	4.09 (3.13)	2.78 (2.33)
GAV		3.14 (2.54)	2.57 (2.29)	1.93 (1.81)
Domain-specific vocabulary	Ecosystem words	1.08 (1.12)	.86 (1.00)	.54 (.79)
	Economy words	.04 (.19)	.08 (.30)	0
	Public policy words	.54 (.67)	.58 (.77)	.31 (.52)
	Total occurrence	1.65 (1.51)	1.52 (1.50)	.86 (1.05)

Table 2. Zero-Order Pearson Correlations Between Use of Academic Vocabulary and Pretest Language and Social Measures ($n = 460$).

	1	2	3	4	5	6	7	8	9	10	11	12
1. Reading comprehension	–											
2. Object naming	.36***	–										
3. Talkativeness	.11*	.10*	–									
4. Good idea nominations	.42***	.13**	.32***	–								
5. Leadership nominations	.33***	.12**	.11*	.71***	–							
6. Indegree centrality	.16***	.04	.18***	.44***	.46***	–						
7. Information centrality	.19***	.13**	.12*	.22***	.22***	.53***	–					
8. Betweenness centrality	.02	.03	.11*	.22***	.23***	.54***	.30***	–				
9. Peer-liking rating	.06	−.04	.12**	.48***	.41***	.60***	.35***	.34***	–			
10. GAV	.21***	.19***	.11*	.09†	.08†	−.00	.12**	.06	−.10*	–		
11. Domain-specific vocabulary	.34***	.20***	.20***	.17***	.14**	.08†	.06	.03	−.02	.41***	–	
12. Total academic vocabulary	.31***	.23***	.17***	.14**	.12**	.03	.11*	.06	−.08†	.91***	.75***	–

†$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$.

condition. Classroom was entered as a random variable. Six covariates—reading comprehension, object naming, talkativeness, peer-liking rating, leadership nominations, and information centrality—were entered sequentially into the analysis. Gender and ethnicity were dropped during model selection because neither of them was ever significant, p's $>.30$. Indegree centrality and betweenness centrality were not included because of low correlations with use of academic words and overlap with information centrality. Good idea nomination was included initially but later dropped because it overlapped with leadership nominations.

The mixed-effects Poisson regression models for the use of GAV are presented in Table 3. The model selection process started with an empty model that included only the random effect. The empty model showed significant individual differences in the use of GAV words, after controlling for classroom variance, intercept = 0.89, $t(36) = 15.05$, $p < .001$. Reading comprehension, a measure of students' receptive language ability, was entered in Model 1 and predicted use of GAV, $F(1, 422) = 15.66$, $p < .001$. Object naming, which is considered to represent students' basic oral English proficiency, was entered in Model 2 and also predicted GAV, $F(1, 421) = 9.46$, $p < .01$. Models 3 through 6 accounted for social characteristics. There were significant effects for talkativeness, $F(1, 420) = 6.70$, $p = .01$; the peer-liking rating of popularity among classmates, $F(1, 419) = 9.90$, $p < .01$; leadership nominations, $F(1, 418) = 4.31$, $p = .039$; and information centrality, a measure of students' status in the classroom social network that takes into account both direct and indirect ties with others, $F(1, 417) = 6.29$, $p = .013$.

After controlling for initial language ability, talkativeness, and social characteristics, we incorporated intervention condition in Model 7 and found a significant effect, $F(2, 417) = 11.73$, $p < .001$. Pair-wise comparisons showed that CG students used more GAV words than DI students, mean difference $M_{diff} = .58$, $t(417) = 2.26$, $p = .024$, and DI students used more GAV words than control students, $M_{diff} = .64$, $t(417) = 2.63$, $p < .01$.

To summarize, students who interacted in CGs productively used more GAV in the whale transfer task than students who experienced DI, controlling for initial language ability, talkativeness, and social characteristics. Both CG and DI students produced more GAV words than uninstructed control students. Students who were better readers and had better oral English, who were more talkative, recognized by peers for leadership, or enjoyed higher social status were more likely to use GAV words in the whale transfer task. The only factor that showed a negative association with GAV words was the peer-liking measure; students who were more liked by classmates tended to use fewer general academic words. The foregoing pattern of results remained the same when we analyzed the total occurrence of academic words including both general and domain-specific academic vocabulary.

Factors predicting use of domain-specific vocabulary. A mixed-effects Poisson regression model was constructed for domain-specific vocabulary with gender, ethnicity, and intervention condition as fixed effects; classroom as a random effect; and the two initial language ability measures and four social

Table 3. Generalized Poisson Mixed Models of Uses of GAV in Whale Interview.

	Empty Model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Fixed effects								
Intercept	.89*** (.06)	.88*** (.06)	.88*** (.06)	.86*** (.06)	1.37*** (.17)	1.47*** (.18)	1.30*** (.19)	1.10*** (.18)
Language ability								
Reading comprehension[a]		.013*** (.003)	.009** (.003)	.008* (.003)	.009** (.003)	.007* (.003)	.008* (.004)	.008* (.004)
Object naming[b]			.008** (.002)	.007** (.003)	.007** (.003)	.007** (.003)	.007** (.003)	.007** (.003)
Social characteristics								
Talkativeness[c]				.27** (.10)	.30** (.10)	.30** (.10)	.26* (.10)	.25** (.10)
Peer-liking[d]					−.17** (.06)	−.23*** (.06)	−.30*** (.07)	−.32*** (.06)
Leadership[e]						.41* (.20)	.39* (.19)	.43* (.19)
Information centrality[f]							3.09* (1.23)	2.95** (1.00)
Intervention condition[g]								
CG vs. DI								.21* (.09)
CG vs. Control								.47*** (.10)
DI vs. Control								.26** (.10)
Random effects								
Variance of intercept	.09** (.03)	.09** (.03)	.09** (.03)	.09** (.03)	.08** (.03)	.08** (.03)	.06** (.02)	.02† (.01)
Fit statistics								
AIC	1,981	1,968	1,961	1,956	1,948	1,946	1,942	1,928

†$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$.
[a] Class-mean-centered Gates-MacGinitie reading comprehension score (corrected for guessing).
[b] Class-mean-centered object naming score (number of objects correctly named per minute).
[c] Rate of classmates' nominations for talkativeness minus rate of nominations for quietness.
[d] Average of classmates' ratings of "how much you like to play with" this individual (five-point Likert scale).
[e] Rate of classmates' nominations for being a "good leader."
[f] Weighted average of direct and indirect friendship ties.

measures as covariates. Both CG and DI students significantly outperformed control students, M_{diff} (CG vs. control) = .79, $t(414) = 4.23$, $p < .001$ and M_{diff} (DI vs. control) = .67, $t(414) = 3.85$, $p < .001$. However, no difference was observed between the CG and DI conditions, $t(414) = .35$, $p = .73$. Girls used more domain-specific vocabulary than boys, $F(1, 414) = 5.04$, $p = .025$. Reading comprehension, object naming, talkativeness, and leadership nominations were significant predictors of use of domain-specific vocabulary, p's $< .05$. Peer-liking ratings showed a marginal negative effect, $F(1, 414) = 3.28$, $p = .071$. Social status (information centrality) did not predict the use of domain-specific vocabulary, $F(1, 414) = .76$, $p = .38$. No difference was found between African American and Latina/o students in use of domain-specific academic words, $F(2, 414) = 1.65$, $p = .19$.

CG, although more effective in improving students' use of GAV, was no better than DI in improving students' domain-specific vocabulary. However, students in the CG condition thoroughly studied only one domain of knowledge, whereas DI students studied all three domains. CG "experts" might use more words from their own domain of expertise as compared with words from the two less studied domains. Ordered logistic regression analyses were conducted to investigate condition differences in use of vocabulary specific to the ecosystem, economy, and public policy domains. The ordinal variable for use of ecosystem words contained four categories: zero frequency ($n = 219$), one type ($n = 148$), two types ($n = 51$), and three or more types ($n = 42$). There were three categories of uses of public policy words: zero frequency ($n = 281$), one type ($n = 143$), and two or more types ($n = 36$). The proportional odds assumption held for both analyses. The control condition was used as the reference. After controlling for gender, ethnicity, reading comprehension, object naming, talkativeness, leader nominations, information centrality, and peer-liking ratings, the condition effect was significant in the use of ecosystem words, χ^2 (4, $n = 460$) = 19.64, $p < .01$, and the use of public policy words, χ^2 (4, $n = 460$) = 16.09, $p < .01$. No analysis of economy words was performed due to their rare occurrence (4%); among the 460 students, 6 CG students and 12 DI students used one economy word.

Students in the public policy group had the highest chance of using ecosystem words (odds ratio Public policy/Control = 3.00, 95% CI [1.49, 6.07]) as compared with the economy group (odds ratio Economy/Control = 2.99, 95% CI [1.48, 6.06]), the ecosystem group (odds ratio Ecosystem/Control = 2.51, 95% CI [1.49, 4.24]), or the DI students (odds ratio DI/Control = 2.08, 95% CI [1.31, 3.30]). Public policy students also had the highest chance of using public policy words (odds ratio Public policy/Control = 4.07, 95% CI [1.92, 8.61]) as compared with the DI students (odds ratio DI/Control = 2.11, 95% CI [1.29, 3.47]). However, both ecosystem and economy students had a lower chance of using public policy words than DI students, odds ratio Ecosystem/Control = 1.63, 95% CI [.92, 2.87] and odds ratio Economy/Control = 1.65, 95% CI [.76, 3.60].

To summarize, students in the three CG expert groups all had a greater likelihood of using ecosystem words than the DI students, and students in public policy groups also used more public policy words than the DI students. Bear in mind, however, that when the distinction among expert groups is collapsed and all domain-specific words are aggregated, there is no overall difference between CG and DI students.

Influence of classroom dialogue on productive use of GAV words

We hypothesized that the use of GAV words during the Wolf Unit would play a part in vocabulary growth and would favor CGs as compared with DI. A simple line of reasoning to inform this hypothesis is that CG provides more opportunities to use GAV words during discussions and other group activities. This, in turn, enhances their ability to productively use these words in the whale transfer interview.

To examine this hypothesis, we performed a mediation analysis at the word level. The analysis encompassed the 158 general academic words used by one or more student during the whale interview. *Control GAV* is defined as the proportion of control students ($n = 147$) who used each of the 158 GAV words in response to the whale question; because control students did not study the Wolf Unit, their use of the words provides the baseline for GAV words already known by a fifth grader from a low-income

minority family that were accessible to this fifth grader while trying to respond to the whale question. *Wolf Print GAV* is the print exposure frequency of each of the array of 158 GAV words in the curriculum materials that CG and DI students were supposed to read during the Wolf Unit. The outcome variable was *Whale GAV*, the proportion of CG students and DI students who used each general academic word in the whale response. *Intervention Condition* (CG = 1, DI = 0) was the explanatory variable of major interest. *Wolf Dialogue GAV*, the frequency of student use of each of the array of general academic words in classroom talk throughout the Wolf Unit, was the candidate mediator variable.

Wolf Dialogue GAV was based on a search for the 158 GAV words in transcripts of 146 4-minute excerpts (total duration = 594 minutes) sampled from 500 hours of video recorded during Wolf Unit lessons in the CG and DI classrooms. The 4-minute excerpts were selected according to a stratified random sampling plan. One excerpt was sampled from each of six important lessons (seven in one case) in each CG classroom ($n = 12$) drawn from the following: introduction to the Wolf Unit, first Collaborative Reasoning discussion, wolves in the United States, wolves and the ecosystem, a poster presentation of major concepts in the ecosystem domain, and the second Collaborative Reasoning discussion. Likewise, one excerpt was sampled from each of six important lessons (seven in one case) in each DI classroom ($n = 12$) to cover the introduction to the Wolf Unit, wolves in the United States, wolves and the ecosystem, wolves and the economy, wolves and public policy, and review of major concepts in the Wolf Unit. The sampled lessons were spaced at approximately equal intervals across the unit and usually occurred on Tuesdays. The first few and last few minutes of each lesson were trimmed because noninstructional activities are likely at these times (passing out materials, lining up for lunch). The 4-minute excerpt was selected at random from the remainder of the lesson with the constraint that in any one classroom three of the six excerpts came from the first half of a sampled lesson and three came from the second half of a sampled lesson.

Separate counts of GAV words were obtained for student uses and teacher uses. The average rate of student use of GAV words per 4-minute excerpt was 4.63 in CG classrooms and 2.15 in DI classrooms. In contrast, the average rate of teacher use of GAV words per 4-minute excerpt was 1.34 in CG classrooms but 3.64 in DI classrooms. In the principal mediation analysis described below, Wolf Dialogue GAV refers to student uses of GAV words during the wolf management unit.

The mediation analysis involved a subset of 88 students from the CG condition and all 153 students from the DI condition who were interviewed about whales. The analysis had to be limited to the subset of CG students whose classroom dialogue had been video recorded. These were the students in ecosystem groups. Their dialogue could have been quite different from the dialogue of students in economy groups or public policy groups who were not video recorded. In DI classrooms, the video recorded dialogue was representative of all classroom dialogue so all 153 DI students who were interviewed about the whale question could be included in the analysis.

Since the outcome variable (Whale GAV) and the mediation variable (Wolf Dialogue GAV) followed a Poisson distribution, we constructed three Poisson regression models to determine the coefficients that define direct and indirect effects. The three Poisson models examined (1) the effect of intervention condition on Whale GAV controlling for Control GAV and Wolf Print GAV (Figure 1A), (2) the effect of intervention condition on Wolf Dialogue GAV after controlling for Wolf Print GAV (Figure 1B), and (3) the effect of intervention condition on Whale GAV after entering the mediator variable Wolf Dialogue GAV and the two covariates, Control GAV and Wolf Print GAV (Figure 1B).

As can be seen in Figure 1B, there are two indirect paths involving Wolf Dialogue GAV. One is *Intervention Condition* → *Wolf Dialogue GAV* → *Whale GAV* and the other is *Wolf Print GAV* → *Wolf Dialogue GAV* → *Whale GAV*. The possible mediating effect of Wolf Dialogue GAV via the path *Intervention Condition* → *Wolf Dialogue GAV* → *Whale GAV* was tested first. The analytical method we used was developed by Iacobucci (2012) for cases in which the independent variable (X), mediator variable (M), and/or outcome variable (Y) are categorical. In line with Baron and Kenny (1986), coefficient a represents the direct effect of X on M, b represents the direct effect of M on Y when X and M are both in the model, c represents the direct effect of X on Y without M in the model,

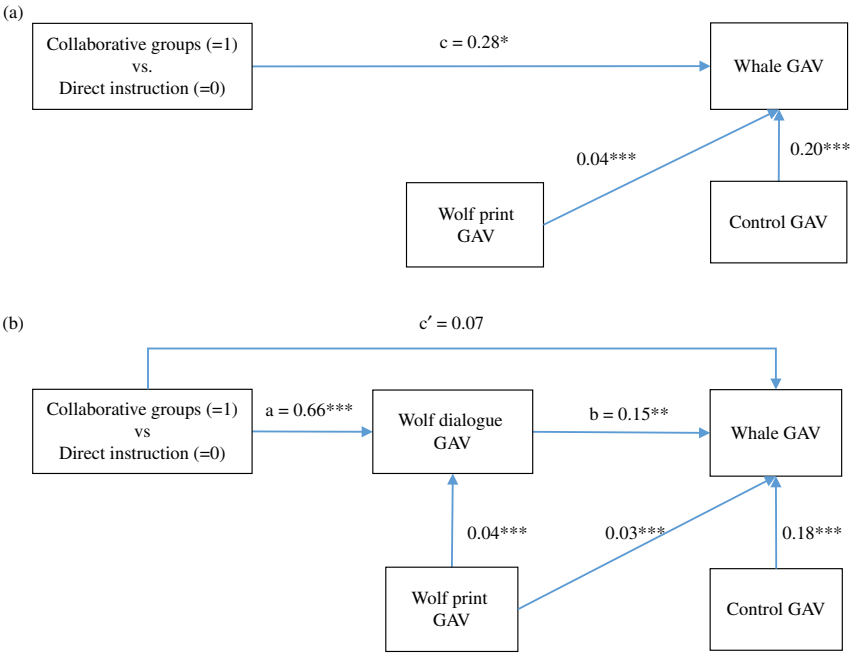

Figure 1. (A) Path model of productive student use of GAV in whale transfer interview, excluding classroom dialogue. (B) Path model of productive student use of GAV in whale transfer interview, including classroom dialogue. *$p < .05$, **$p < .01$, ***$p < .001$.

and c' the effect of X on Y with M in the model. The standard errors of a and b are denoted as s_a and s_a and can be obtained from the regression analyses. The test statistic $Z_{mediation}$ was defined by Iacobucci (2012) as

$$Z_{mediation} = \frac{\frac{a}{s_a} \times \frac{b}{s_b}}{\sqrt{(\frac{a}{s_a})^2 + (\frac{b}{s_b})^2 + 1}}$$

Poisson Model 1 establishes that intervention condition significantly predicted Whale GAV, $c = .28$, χ^2 $(n = 316) = 4.90$, $p = .027$, after controlling for Control GAV and Wolf Print GAV. Poisson Model 2 indicates a significant effect of intervention condition on Wolf Dialogue GAV, $a = .66$, $\chi^2(n = 316) = 30.43$, $p < .001$. In Model 3 the direct effect of intervention condition on Whale GAV was not significant after incorporating the mediator variable Wolf Dialogue GAV in the model, $c' = .07$, χ^2 $(n = 316) = .23$, $p = .64$. Instead, in Model 3 there is a significant effect of Wolf Dialogue GAV on Whale GAV, $b = .15$, χ^2 $(n = 316) = 8.17$, $p < .01$. Based on Iacobucci's (2012) method, there is a mediation effect of Wolf Dialogue GAV, $Z_{mediation} = 2.51$, $p < .01$. Therefore, the analysis indicates that Wolf Dialogue GAV mediated the effect of intervention condition on Whale GAV.

We also tested whether there was a mediation effect from teachers' use of GAV words in classroom dialogue during the wolf management unit. The results indicated that teachers' use of GAV words did not predict students' use of GAV words in the whale interview and did not mediate the effect of instructional condition on Whale GAV: $a = -.50$, χ^2 $(n = 316) = 12.75$, $p < .001$; $b = .05$, χ^2 $(n = 316) = .85$, $p = .36$; $c = .28$, χ^2 $(n = 316) = 4.90$, $p = .027$; $c' = .31$, χ^2 $(n = 316) = 5.63$, $p = .018$; and $Z_{mediation} = -.86$, $p = .19$. An analysis of total GAV words in classroom dialogue that included both teacher and student uses of words yielded results similar to the analysis involving only student uses, but the mediation effect of Wolf Dialogue GAV was not as strong.

The second possible mediating effect of Wolf Dialogue GAV is via the path *Wolf Print GAV → Wolf Dialogue GAV → Whale GAV*. The analysis shows that Wolf Dialogue GAV partially mediated the effect of Wolf Print GAV on Whale GAV, $Z_{mediation} = 2.75$, $p < .01$. To summarize, the frequency with

which students used GAV in classroom dialogue mediated the effect of intervention condition and partially mediated the influence of the printed wolf management curriculum on use of GAV words in the whale interview.

Discussion

An important finding from the present research is that students who studied wolf management for 6 weeks, addressing the question of whether a community should be permitted to eradicate a pack of wolves, exceeded uninstructed control students in productive use of academic vocabulary in an open-ended transfer task that required students to justify a position on whether or not whaling should be allowed. Students who experienced the Wolf Unit exceeded control students in use of both GAV words, such as *increase* and *protect*, and domain-specific vocabulary words, or technical terms, such as *population* and *extinct*.

Few if any previous studies have demonstrated that instruction can have a broad impact on students' ability to productively use academic words, where by *productive use* we mean the ability to say or write the words spontaneously, in contrast to recognizing the words, selecting their definitions from a set of options, or supplying them in response to focused prompts. That students had active control of a cross section of the academic vocabulary words they encountered in the Wolf Unit is further suggested by the fact that they were able to use the words in response to the whale question, which differed in many surface aspects from the wolf question they had answered during the unit.

Our general explanation for why students used more academic vocabulary words in the oral transfer task after completing the Wolf Unit is that the unit provided for multiple encounters with words in meaningful contexts. Students were exposed to a wide variety of academic words as they studied different topics related to the controversial policy issue about wolves. To make a thoughtful decision about the policy issue, students needed to organize information about the different topics into arguments. Students in CGs acquired academic words useful for organizing and expressing information about wolves in give and take with peers, whereas students who received DI acquired the words by following the teacher's explanations and using the words in response to the teacher's questions.

A second important finding of this research is that students who experienced the Wolf Unit via CGs used more GAV words on the oral transfer task than students who experienced the Wolf Unit via DI. Classroom dialogue during collaborative peer interaction and teacher-led instruction differs in many ways (Chinn, Anderson, & Waggoner, 2001), and it is here that we looked for an explanation for the greater use of GAV words by CG than DI students in response to the whale question. A simple explanation for the difference is that GAV words were used more frequently in the classroom dialogue in CG classrooms.

This simple explanation was evaluated in the analysis summarized in Figure 1. After controlling for prior knowledge (Control GAV) of an array of 158 GAV words that appeared in the whale interviews, CG students were significantly more likely than DI students to use the array of words in response to the whale question (Whale GAV). But, after incorporating the frequency of students' use of the words in classroom talk in CG and DI classrooms (Wolf Dialogue GAV) in the model, the direct effect of instructional condition was no longer significant. In this model, the effect of instructional condition was indirect. Instructional condition predicted student use of GAV words in classroom dialogue and student use of the words in classroom dialogue predicted their use in the whale interview. Thus, the effect of CG appears to have been mediated at least in part by the greater frequency of academic vocabulary words in students' talk in CG classrooms.

Beyond frequency of word use, another factor that could have contributed to the advantage of CG students is heightened engagement. Wu et al. (2013) reported that collaborative peer discussions result in greater student interest and engagement than conventional teacher-led discussions. Students motivated to actively join a discussion may thereby use some academic vocabulary words and more closely follow others' use of academic words.

More opportunity to *actively use* academic vocabulary words in dialogue figures to be a major reason for CG students' greater uptake and later use of the words. Opportunities for speaking are expanded when a class is divided into small groups. Time must be divided among all the students in whole-class discussions whereas time is divided among fewer students in small-group discussions. During teacher-led discussions, time must be split between the teacher and the students, and teachers often take much of the time and may express more than half the words that are spoken during discussions (Cazden, 2001). In collaborative small-group discussions among peers, teacher time is nil and almost all of the time is available for student speaking turns. Dispensing with the rigmarole of hand raising and teacher nomination for speaking turns saves additional time during peer-to-peer discussions and enables a focus on ideas. It is not surprising, therefore, that Chinn, Anderson, and Waggoner (2001) found that student words per minute was nearly twice as high in peer-managed discussions in small groups as compared with teacher-led discussions in the same small groups. In the present study, the rate of use of GAV words during classroom dialogue by CG students was over twice as high as the rate of use by DI students. Conversely, the rate of use was nearly three times higher among DI teachers than CG teachers.

Richer classroom dialogue is a probable factor contributing to the advantage of CG students as compared with DI students in use of GAV words in the whale interview. A body of evidence establishes that improved comprehension, learning, and problem solving are associated with high quality classroom discussion during which students' provide explanations (as opposed to merely listening to them), elaborate ideas by linking them with prior knowledge, predict the consequences of different courses of action, draw inferences that connect different parts of texts, construct analogies between real and imagined situations, consider alternative explanations, support ideas with evidence, build on one another's ideas to co-construct explanations, and critique one another's ideas (Murphy et al., 2009; Nystrand et al., 1997; Resnick & Schantz, 2015; Reznitskaya et al., 2009). The link to vocabulary acquisition comes from a large study by Lawrence, Crosson, Paré-Blagoev, and Snow (2015) encompassing over 1,500 middle school students from 28 schools. That study evaluated a vocabulary instruction program called Word Generation, in which students read, discuss, and write about controversial topics using a list of target academic words. The results showed that the Word Generation program improved the quality of discussion in a range of classes including math, science, language arts, and social studies. The program also led to modest gains in academic vocabulary. These gains were mediated in part by quality of classroom discussion.

A low-inference indicator of the quality of discussions in CG and DI classrooms is use of coordinating conjunctions. Morris et al. (2013) examined the frequency of use of the conjunctions *because, so, if, then, and*, and *but* during the 4-minute excerpts sampled from the classroom dialogue in the CG and DI classrooms enrolled in the present study. CG students' rate of use of the conjunctions was four times higher than the rate of use by DI students. In contrast DI teachers used the conjunctions at twice the rate of CG teachers. Thus, CG students had the experience of expressing elaborated and connected ideas during CG. DI students, in contrast, depended on teachers to initiate ideas and make connections.

High quality discussion may lead to better lexical representations. According to the Lexical Quality Hypothesis (Perfetti & Hart, 2002), a good lexical representation contains sufficient knowledge of orthographic, phonological, and semantic properties of words for people to retrieve the words rapidly and flexibly. Students who experience the Wolf Unit may develop high quality lexical representations of academic words that entail a network of connections that link the words to many contexts. Students in CGs, especially, may develop dense and integrated networks through constantly presenting claims and responding to classmates' claims and challenges during collaborative discussions. The more dense and integrated the network the more likely words are to "come to mind" in new contexts such as the whale interview.

The wolf management curriculum is encapsulated in printed materials that provide information, explain concepts, and introduce perspectives on the relationships between wolves and the ecosystem, wolves and the economy, and wolves and public policy. However, words lying lifeless on a page do not teach. The curriculum is brought to life in classroom dialogue. The analysis summarized in Figure 1

shows that the frequency of GAV words in printed Wolf Unit curriculum materials (Wolf Print GAV) influenced the frequency of these words in classroom dialogue (Wolf Dialogue GAV) which in turn influenced whether students would use the words in response to the question about whaling (Whale GAV). Therefore, in other words, the influence of the printed curriculum on use of academic vocabulary in the whale interview was mediated by classroom dialogue.

On average, CG students used two more academic word types in the whale interview than control students, whereas DI students used about one-and-a-third more types than control students. These perhaps seem like minor improvements, but it should be stressed again that the children were from underserved communities and more than half were English language learners, and thus probably most had little exposure to academic language. The words observed in any finite sample of language are a fraction of the lexicon from which the words were drawn (see Carroll, 1971; Kojima & Yamashita, 2014). For every additional word observed in the language sample it must be inferred that other words added to the lexicon that remained unobserved. If a 6-week-long unit can result in one or two more academic words appearing in a limited sample of speech, other words not observed must also have been acquired and the long-term yield of new words from this kind of instruction could be substantial.

Talkativeness, nominations for leadership, and position in the classroom social network (information centrality) were positively related to use of GAV words in the whale transfer interview, whereas peer-liking showed a negative relationship. These findings provide tantalizing clues to what must have been the social process that gave rise to academic vocabulary knowledge. We surmise that talkative, socially centered children took the lead in using academic vocabulary words and trying to figure out their meanings. In a study of peer influences during collaborative reasoning, Lin et al. (2015) concluded that, "students at the center of the classroom friendship network play an influential role in creating a stimulating and friendly discussion environment in which everyone has the opportunity to make contributions" (p. 94). The negative relationship between peer-liking and use of academic vocabulary apparently means that, aside from children who are well-liked because of their intellectual and social leadership, the remaining children who are well-liked tend to be averse to academic talk and to shy away from using academic vocabulary words.

CG students and the DI students did not differ overall in use of domain-specific vocabulary in the whale interview. However, when the fact that "expert" groups within the CG condition concentrated on different domains of knowledge is taken into consideration, a somewhat different picture emerges. Students in ecosystem expert groups more often used ecosystem words than DI students, and students in public policy expert groups more often used public policy words than DI students. Results were flat for economy words because these were seldom used in the whale interview. Interestingly, even students in the public policy and economy expert groups used significantly more ecosystem words than DI students. With the caveat that we have not yet examined exactly which words were taught and how they were taught, the findings so far seem to give reason to doubt the received wisdom that technical vocabulary cannot be learned unless explicitly taught.

A limitation of this study is that there was no preintervention assessment of children's ability to productively use academic vocabulary in an open-ended speaking task. A preintervention assessment would have provided further assurance that groups were comparable and allowed more sensitive tests of intervention effects. A second limitation is that untangling the effects of instructional conditions on domain-specific vocabulary acquisition was tricky because of the complicated instructional design. CG students studied one domain of knowledge intensively, whereas DI students were exposed to all three domains. A third limitation was that because of constraints on time and resources, only 60% of the students could be interviewed about the whale question and the analysis of classroom dialogue was limited to 2% of the available lesson video.

Although productive vocabulary has long been distinguished from receptive vocabulary, "the idea of productive vocabulary remains a fundamentally elusive one. The main reason for this is that it has proved surprisingly difficult to develop simple and elegant tests of productive vocabulary . . . " (Meara & Alcoy, 2010, p. 222). Laufer and Nation (1999) developed a productive vocabulary assessment that

may meet the requirement of being simple and elegant. It entails completing words in sentences. The first letters of each incomplete word are provided. The letter cues and cues from sentence meaning converge on one and only one word, so although the method assesses productive vocabulary, it does so only at a minimal level.

As far as we know, the present study is the first to demonstrate acquisition and spontaneous productive use of academic vocabulary in an open-ended oral transfer task. Most research on vocabulary instruction employs tests of receptive knowledge of words explicitly taught. A standardized vocabulary test containing mostly words beyond those taught may be given in addition, in the hope that students have learned something generalizable or transferable. Usually this is a vain hope. Vocabulary instruction seldom improves performance on standardized vocabulary tests, perhaps because such tests do not fully reflect students' vocabulary competence (cf. Pearson, Hiebert, & Kamil, 2007). Findings of the present study warrant renewed optimism that instruction can have a broad impact on students' ability to understand and productively use academic vocabulary words.

Funding

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, Principle 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 Matthew Sutton, Shengcheng Yuan, Aini Marina Ma'rof, Nikisha Blackmon, and Xiaoying Wu.

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, NY: Guilford.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

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.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding. *American Educational Research Journal*, *40*, 685–730.

August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. J. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness*, *2*, 345–376.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice*, *20*, 50–57.

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

Barron, R. F., & Melnick, R. (1973). The effects of discussion upon learning vocabulary meanings and relationships in tenth grade biology. In H. L. Herber & R. F. Barron (Eds.), *Research in reading in the content areas: Second year report* (pp. 46–52). Syracuse, NY: Syracuse University, Reading and Language Arts Center.

Bailey, A. L. (Ed.). (2006). *Language demands of school: Putting academic English to the test*. New Haven, CT: Yale University Press.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford.

Bravo, M. A., & Cervetti, G. N. (2008). Teaching vocabulary through text and experience in content areas. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about vocabulary instruction* (pp. 130–149). Newark, DE: International Reading Association.

Butts, C. T. (2008). Social network analysis with sna. *Journal of Statistical Software*, *24*, 1–51.

Carlisle, J. F., Fleming, J. E., & Gudbrandsen, B. (2000). Incidental word learning in science classes. *Contemporary Educational Psychology*, *25*, 184–211.

Carlo, M., August, D., McLaughlin, B., Snow, C., Dressler, C., Lippman, D., … White, C. (2004). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, *39*, 188–215.

Carroll, J. B. (1971). Statistical analysis of the corpus. In J. B. Carroll, P. Davies, & B. Richman (Eds.), *American Heritage Word Frequency Book* (pp. xxi–xl). Boston, MA: Houghton Mifflin.

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

Corson, D. (1997). The learning and use of academic English words. *Language Learning*, *47*, 671–718.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213–238.

Ellis, R. (2005). *Instructed second language acquisition: A literature review*. Wellington, New Zealand: Ministry of Education.

Garcia, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic students. *Reading Research Quarterly*, *26*, 371–392.

Gillies, R. M. (2007). *Cooperative learning: integrating theory and practice*. Thousand Oaks, CA: Sage.

Hiebert, E. H., & Lubliner, S. (2008). The nature, learning, and instruction of general academic vocabulary. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about vocabulary instruction* (pp. 106–129). Newark, DE: International Reading Association.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*, 55–88.

Howe, C., & Mercer, N. (2007). *Children's social development, peer interaction and classroom learning* (Primary Review Research Survey 2/1b). Cambridge, UK: : University of Cambridge.

Iacobucci, D. (2012). Mediation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology*, *22*, 582–594.

Institute of Educational Sciences. (2007). *Organizing instruction and study to improve student learning*. Washington, DC: IES.

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 & L. Kucan (Eds.), *Bringing reading research to life: Essays in honor of Isabel L. Beck* (pp. 170–193). New York, NY: Guilford Press.

Keller, K., Troesch, L. M., & Grob, A. (2015). A large receptive-expressive gap in bilingual children. *Frontiers in Psychology*, *6*, 1284.

Kelley, J. G., Lesaux, N. K., Kieffer, M. J., & Faller, S. E. (2010). Effective academic vocabulary instruction in the urban middle school. *Reading Teacher*, *64*, 5–14.

Kojima, M., & Yamashita, J. (2014). Reliability of lexical richness measures based on word lists in short second language productions. *System*, *42*, 23–33.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*, 307–322.

Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language Testing*, *16*, 33–51.

Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., & Snow, C. E. (2015). Word generation randomized trial: discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal*, *52*, 750–786.

Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly*, *45*, 196–228.

Lin, T.-J., Anderson, R. C., Hummel, J. E., Jadallah, M., Miller, B. W., Nguyen-Jahiel, K., … 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., … Li, Y. (2015). Social influences on children's development of relational thinking during small-group discussions. *Contemporary Educational Psychology*, *41*, 83–97.

Ma, S., Anderson, R. C., Lin, T.-J., Zhang, J., Morris, J. A., Nguyen-Jahiel, K., … Latawiec, B. (2015). *Instructional influences on English language learners' storytelling*. Champaign, IL: Center for the Study of Reading.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *80*, 300–335.

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.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, *20*, 522–535.

Meara, P. M., & Alcoy, J. E. O. (2010). Words as species: An alternative approach to estimating productive vocabulary size. *Reading in a Foreign Language*, *22*, 222–236.

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

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, *79*, 979–1007.

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, S., & Faulkner, D. (2000). Learning to collaborate: Can young children develop better communication strategies through collaboration with a more popular peer. *European Journal of Psychology of Education*, *15*, 389–404.

Nagy, W. E. (1997). On the role of context in first- and second-language vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 64–83). Cambridge, UK: Cambridge University Press.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *19*, 304–330.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*, 237–270.

Nagy, W. E., & Scott, J. (2000). Vocabulary processes. In M. Kamil, P. Mossenthal, P. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 269–284). Mahwah, NJ: Erlbaum.

Newcomb, A. F., Bukowski, W. M., & Pattee, L. (1993). Children's peer relations: a meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin*, *113*, 99.

Nystrand, M., Gamoran, A., Kachur, R., & Pendergast, C. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York, NY: Teachers College Press.

Oxford, R. L. (1997). Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. *Modern Language Journal*, *81*, 443–456.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, *42*, 282–296.

Resnick, L. B., & Schantz, F. (2015). Talking to learn: The promise and challenge of dialogic teaching. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 441–450). Washington, DC: American Educational Research Association.

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.

Salahu-Din, D., Persky, H., & Miller, J. (2008). *The nation's report card: writing 2007* (NCES 2008-468). Washington, DC: National Center for Educational Statistics. Institute for Educational Sciences, U.S. Department of Education.

Scarcella, R. (2003). *Academic English: A conceptual framework*. Los Angeles, CA: University of California Language Minority Research Institute.

Shoebottom. (2008). General academic vocabulary [web log post]. Retrieved from http://esl.fis.edu/vocab/index-a.htm

Schwanenflugel, P. J., Stahl, S. A., & McFalls, E. L. (1997). Partial word knowledge and vocabulary growth during reading comprehension. *Journal of Literacy Research*, *29*, 531–553.

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., Lawrence, J., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, *2*, 325–344.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, *328*, 450–452.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *Cambridge handbook of literacy*. New York, NY: Cambridge University Press.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, *56*, 72–110.

Stahl, S. A., & Vancil, S. J. (1986). Discussion is what makes semantic maps work in vocabulary instruction. *Reading Teacher*, *40*, 62–69.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–407.

Townsend, D., & Collins, P. (2009). Academic vocabulary and middle school English learners: An intervention study. *Reading and Writing*, *22*, 993–1020.

Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, *18*, 137.

Vaughn, S., Martinez, L. R., Linan-Thompson, S., Reutebuch, C. K., Carlson, C. D., & Francis, D. J. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness*, *2*, 297–324.

Wu, X., Anderson, R. C., Nguyen-Jahiel, K., & Miller, B. (2013). Enhancing motivation and engagement through collaborative discussion. *Journal of Educational Psychology*, *105*, 622–631.

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

Appendix A

Whale transfer interview

Hunting whales is called whaling. People have been hunting whales for over a thousand years, but now people have different opinions about whaling.

Some people say that we should hunt whales, because they are eating too many fish. Whales are the largest animals living in the oceans, and they need to eat a huge amount of fish and other sea creatures every day. For example, some whales may eat more than 8,000 pounds of food a day. Several kinds of fish that whales eat are already disappearing, and only a few of them are left. Other people do not think that whales are eating too much fish. These people argue that whales are not the only animals that eat fish. Most of the fish are eaten by people, by other fish, and by seabirds. In fact, some kinds of whales do not eat fish at all. Instead, they eat very small plants and other tiny animals.

Whales also have an effect on the businesses of different people. People who hunt whales argue that whaling provides jobs to them and the people who work in restaurants. Whale hunters can make a lot of money selling whales to restaurants. The meat from one whale can feed as many as 1,000 people for almost two months. In some countries, like Norway, whale meat is a major source of food. Other people say that whaling hurts whale-watching businesses. Each year, millions of people take tours to watch whales in the ocean. These people spend money on boats, travel, hotels, and food. Whale-watching makes a lot of money for people in many countries.

Some people are worried that whaling will affect whale populations. These people say that there are not many whales left in the ocean, because too many whales have been killed by hunters and there is less food for the whales to eat. They say that if we keep hunting whales, whales could disappear forever. However, people who want whaling argue that not all kinds of whales are endangered. Some kinds of whales have always been plentiful. Other kinds of whales were few in number in the past, but their numbers are now increasing. Also, in countries like Norway and Japan, whaling is a tradition. Keeping this tradition is very important for people in these countries.

Big question

Do you think we should allow people to hunt whales?

Prompt 1: Only if reasons are omitted

1. Tell me why you think we should [should not] allow people to hunt whales.

Prompt 2: Only if counter-argument is omitted

1. Could there be people who do not agree with you?
2. What would be their opinion and reasons?

Prompt 3: Only if rebuttal is omitted

1. What would you say to people who do not agree with your position?

Appendix B

Print Exposure of Domain-Specific Vocabulary in the Wolf Unit.

Ecosystem Words ($n = 25$)	CG Ecosystem Booklets	DI Booklets	Economy Words ($n = 17$)	CG Economy Booklets	DI Booklets	Public Policy Words ($n = 18$)	CG Public Policy Booklets	DI Booklets
balance	12	7	agriculture	6	6	advocate	11	7
carnivore	9	7	compensation	5	6	biologist	12	8
competitor	3	4	compete	5	7	citizen	18	18
consumer	30	12	cost	5	5	common good	30	17
ecosystem	133	60	economy	68	30	community	23	20
endangered	6	11	expense	7	7	conflict	3	3
endurance	2	2	income	12	12	culture	2	2
extinct	5	4	livestock	41	35	heritage	2	2
food web	46	21	logger	20	7	interest(s)	22	26
global warming	4	5	lumber	6	5	majority	26	19
habitat	3	8	manufacturing	4	4	minority	26	19
herbivore	6	4	permit (noun)	6	6	need(s) (noun)	1	1
nature	60	17	profit	4	4	opinion	36	20
natural resource	4	18	ranching	90	74	position	31	14
omnivore	6	4	service	6	6	public policy	33	14
population	33	26	timber	48	26	represent	17	11
predator	23	24	tourism	23	20	right(s) (noun)	17	15
prey	26	21				tradition	2	2
producer	26	12						
recover	3	3						
reintroduce	26	12						
scavenger	6	7						
species	41	6						
starvation	1	2						
territory	12	12						

Appendix C

First 50 Most Frequent General Academic Words in the Wolf Unit.

reason	increase	passage	except	major
agree	decrease	source	experience	hardly
hurt	lose	deserve	protect	professional
affect	survive	worth	destroy	supply
instead	since	blame	regular	surface
business	attack	common	action	belong
disagree	Amount	example	hire	choice
generate	Environment	basic	law	continue
spend	Fault	provide	produce	disturb
argument	Harm	decision	exist	earn