Alberta Journal of Educational Research, Vol. 61.1, Spring 2015, 40-64

The Effects of Discussion and Persuasion Writing Goals on Reasoning, Cognitive Load, and Learning

Perry D. Klein, Jacqueline S. Ehrhardt

Western University

Argumentation can contribute significantly to content area learning. Recent research has raised questions about the effects of discussion (deliberation) goals versus persuasion (disputation) goals on reasoning and learning. This is the first study to compare the effects of these writing goals on individual writing to learn. Grade 7 and 8 students learned about buoyancy through argument writing. A 2 x 2 x 2 between-subjects pretest-post-test randomized experiment was used to investigate the effects of two types of argument writing goals (persuasion versus discussion), two distributions of writing subgoals (segmented versus clustered), and two levels of writing achievement (low versus high) on bias/balance in reasoning, cognitive load, and learning. Results showed that segmented subgoals were rated less difficult than clustered subgoals. In a three-way interaction, for high-achieving writers, subgoal segmentation reduced cognitive load in discussion writing, but increased it in persuasive writing. Argument goal type and subgoal distribution affected bias/balance in claims and inferences. These results suggest that the effects of argument goal type are moderated by subgoal distribution and previous writing achievement.

L'argumentation peut contribuer de façon significative à l'apprentissage dans les cours à contenu. Des recherches récentes ont soulevé des questions quant aux effets des objectifs de la discussion (délibération) par rapport aux objectifs de la persuasion (contestation) sur le raisonnement et l'apprentissage. Cette étude est la première à comparer les effets de ces objectifs de rédaction sur des élèves apprenant à écrire. Des élèves en 7e et 8e année ont appris au sujet de la flottabilité en écrivant des textes argumentatifs. Une expérience 2 x 2 x 2 inter-sujet avant-après avec randomisation a porté sur les effets de deux types d'objectifs de textes argumentatifs (persuasion ou discussion), deux distributions de sous-objectifs de rédaction (segmentation ou regroupement) et deux niveaux de rendement en écriture (bas ou élevé) sur l'équilibre quant au raisonnement, la charge cognitive et l'apprentissage. Les résultats ont indiqué que les sous-objectifs segmentés étaient perçus comme étant moins difficiles que les sous-objectifs regroupés. Dans une interaction à trois, la segmentation des sous-objectifs a réduit la charge cognitive dans les rédactions explicatives, mais l'a augmentée dans les textes argumentatifs chez les écrivains de haute qualité. Le type d'objectifs et la distribution des sous-objectifs ont affecté l'équilibre dans les affirmations et les inférences. Ces résultats portent à croire que les effets du type d'objectifs sont modérés par la distribution des sous-objectifs et par le rendement antérieur en rédaction.

The past three decades of research have shown that writing can contribute significantly to learning in content area subjects (Bangert-Drowns, Hurley, & Wilkinson, 2004; Hebert, Gillespie, & Graham, 2013; Nückles, Hübner, & Renkl, 2009). One genre of writing that contributes to learning is argumentation (e.g., Langer & Applebee, 1987; Wiley & Voss, 1999). Argument writing and instruction in argument writing have been found to contribute to learning in domains such as literature (Kieft, Rijlaarsdam, & van den Bergh, 2008; Lewis & Ferretti, 2011) and history (De La Paz & Felton, 2010; Wiley & Voss, 1999).

Argumentation is also important in science education, a point that has been made from several different theoretical viewpoints for multiple reasons (Osborne, 2010; Schwarz, 2009). With respect to the nature of science, opportunities for argumentation illustrate that science is a social activity; knowledge is developed through discourse, and evaluated through critical debate with reference to evidence, rather than derived from authority (Cavagnetto, 2010; Kuhn, 2010; Osborne, 2010). Science education is an opportunity for students to develop argumentation skills, an approach which has been characterized as "learning to argue" (Kuhn, 2010; Osborne, Erduran, & Simon, 2004; Schwarz, 2009; Zohar & Nemet, 2002). This includes understanding the nature of evidence in science (Osborne, 2010; Schwarz, 2009). Argumentation is also part of the process through which science informs debate concerning controversial social issues (e.g., Zohar & Nemet, 2002).

Argumentation is of interest in the present study primarily because it allows students to reason critically about the relationship between specific scientific theories and evidence (Kuhn, 1993, 2010; Osborne, 2010). Students hold misconceptions about many topics; argumentation allows them to consider contrasting explanations and to evaluate them on the basis of evidence, contributing to conceptual change, a form of "arguing to learn" (Asterhan & Schwarz, 2007; Nussbaum & Sinatra, 2003; Schwarz, 2009). Scientific argumentation also familiarizes students with the evidence for specific theories (Klein & Samuels, 2010). During the past 20 years, in order to support these multiple roles for argumentation—demonstrating the nature of science, learning to argue, and arguing to learn—authors have repeatedly called for increased research on argumentation in science education (Kuhn, 1993; Osborne et al., 2004; Osborne, 2010). Consequently, several innovative science education projects have foregrounded argumentation (Hand, Wallace, & Chang, 2004; Kuhn, 2010; McNeill & Krajcik, 2009; Osborne, et al., 2004; Sandoval & Reiser, 2004; see Rapanta, Garcia-Mila, & Gilabert, 2013, for a review).

To date, most studies of argumentation in science have involved dialogical argumentation through oral discussion or online interaction, rather than individual writing (e.g., Asterhan & Schwarz, 2007; Nussbaum & Sinatra, 2003; Rapanta et al., 2013; Yeh & She, 2010). However, individual writing is an important classroom activity; e.g., in the provincial curriculum by the Ministry of Education for the Province of Ontario (2006) and the Common Core State Standards (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). Consequently, further research is needed concerning the role of argument writing in science.

Two Families of Argumentation Goals

In considering argumentation for learning, a fundamental question is what type of goal students should pursue (Felton, Garcia-Mila, & Gilabert, 2009; Kroll, 2005; Nussbaum, 2005, 2008a). Traditionally, students have been assigned the goal of *persuading* a reader; this kind of writing has been referred to as "disputation" or "opinion" writing. To pursue a persuasion goal, the writer states a claim and provides evidence, addresses counter-arguments, and rebuts alternative claims and evidence. Persuasive argumentation has been employed in many previous studies, instructional resources and curriculum guidelines (e.g., Duke, Caughlan, Juzwik, & Martin, 2012; Ferretti, Lewis, & Andrews-Wecherly, 2009; National Governors Association, 2010). However, questions have recently been raised about the effects of persuasion goals on discourse, reasoning, and learning. Prima facie, persuasion invites the student to defend her or his opinion, rather than to question it or entertain alternative claims (e.g., Felton et al., 2009; Kroll, 2005; Rapanta et al., 2013). In this respect, persuasive goals appear have the potential to evoke students' "myside" bias, a tendency to ignore or exclude evidence that does not support the writer's own view (Wolfe, Britt, & Butler, 2009). This issue could be particularly problematic in science argumentation, because the student's initial view could comprise a misconception, which the student needs to examine critically.

As an alternative to persuasion, researchers have recently begun to investigate a second family of argumentation goals, including deliberation (Felton et al., 2009), exploration (Mercer, 2000), and reflection (Nussbaum, 2008a). These goals require students to develop their own claim rather than to defend a pre-existing claim, and to consider alternatives, rather than to undermine them. Several studies have now shown that goals in the deliberation family elicit a greater variety of argument moves than goals in the persuasion family (Felton et al., 2013; Keefer, Zeitz, & Resnick, 2000; Nussbaum, 2005; Nussbaum & Kardash, 2005). For example, a recent study on reasoning about a science-based social issue showed that in dyadic discussions, the goal to come to a consensus, compared to the goal to persuade, led to a higher level of argument quality with a wider variety of argumentation moves, including more rebuttals (Garcia-Mila, Gilabert, Erduran, & Felton, 2013).

Previous studies of persuasion and discussion have focused on social issues, or social controversies related to science. The controversial, complex, and subjective nature of these issues has made it unclear whether these two types of argument goals produce objectively different effects on reasoning and learning. Consequently, most studies of argument goals have incorporated dependent measures comprised of discourse characteristics, rather than content-based measures of reasoning and learning. In this study, we used the well-defined problem space of buoyancy to investigate the effects of argument writing goals on reasoning and learning. We define reasoning in this context as the coordination of claims with empirical evidence (Kuhn, 2010; Nussbaum, Sinatra, & Poliquin, 2008; Schwarz, 2009).

We consider argumentative reasoning to include three aspects: claims, evidence, and inference. Each can be defined on a spectrum from *biased* to *balanced*. *Claim identification* can range from biased, in which only the claim of the writer is presented, to balanced, in which one or more alternative claims is also identified. *Evidence selection* can range from biased, in which a narrow range of evidence is discussed, to balanced, in which a wide range of evidence is discussed. Particularly important in this process is the writer's inclusion of critical evidence that distinguishes between possible claims. The third aspect, *inference*, can range from myside inference bias, in which a writer presents one claim based on logically insufficient evidence, to balanced inference, in which the writer explicitly and validly relates both her or his own claim and an alternative claim to evidence. Note that for each of these dimensions, balance does not require that the writer agree with a claim different from her or his own; balance only requires reasoned discussion of different claims.

In this study, we extend research on the effects of argument goals to their effects on writing and learning. The only previous study that investigated the effects of argument goals on learning was conducted by Felton et al. (2009). The researchers assigned students to engage in one of the following: (a) oral deliberation to reach consensus with a peer holding a different view; (b) oral disputation to persuade a peer holding a different view; or (c) a control condition. Students in all three conditions then individually wrote an argument text. Deliberation goals resulted in significantly greater content learning than the control condition, while disputation goals did not; however, the effect of deliberation on learning did not differ significantly from that of persuasion. Whereas previous studies have focused primarily on oral discussion (e.g., Felton et al., 2013; Garcia-Mila et al., 2013), the present study investigated argument goals in the context of individual writing. The first research question was the following: In individual writing, would a discussion goal lead to more balanced reasoning and greater learning than a persuasion goal?

Argument Goals and Cognitive Load

Writing involves multiple processes, which impose a substantial load on working memory (Hayes & Chenoweth, 2006; McCutchen, Teske, & Bankston, 2008; Nussbaum, 2008b; Olive, Kellogg, & Piolat, 2008; Shehab, 2011). Learning activities that attempt to manage working memory load are investigated in the context of cognitive load theory (e.g., Sweller et al., 2011). A recurring finding is that the goal structure of tasks affects cognitive load (Ayres, 2006; Kester, Kirschner, & van Merriënboer, 2006). In a classic study of problem solving in mathematics, Sweller (1988) found that tasks that required students to set goals and use means-end problem solving to set subgoals, led to increased cognitive load and decreased learning, relative to a condition that did not require means-end problem solving. More generally, the cognitive load literature shows that when learners must consider several interacting elements of a problem, rather than considering fewer interacting elements or considering elements sequentially, then cognitive load is increased and learning is reduced, particularly for students lower in previous knowledge (see Sweller et al., 2011, for a review).

Recently, researchers have begun investigating the cognitive load of writing activities (Nückles, et al., 2010; Shehab, 2011; Zhu & Zhang, 2005). This raises the question, would argument goal type affect cognitive load? In persuasive writing, the student is expected to state a claim at the beginning of the text. Because elementary school writers typically do not spontaneously plan prior to writing, the student may need to consider at least two competing claims, and the evidence or against each, *before* choosing a claim. It was hypothesized that this would impose a substantial load on working memory. Conversely, in discussion writing, the student is not required to form a claim before writing. Instead, the student may write about one claim and the evidence for or against it, and then write about a second claim and the evidence for or against it. Then, in the conclusion the student could decide which claim is better supported. This reduces the need to look ahead and evaluate the evidence on both sides of an issue, before beginning to write. This raised the question of whether the type of argument goal would affect cognitive load.

Subgoal Prompting

In order to pursue writing goals such as deliberation or persuasion, writers elaborate them through subgoals (Ferretti et al., 2009; Hayes, 2012). The writer may elaborate the goal of persuasion by setting subgoals to do the following: state a claim; present evidence for the claim; state an alternative claim and the reasons for it; and present a rebuttal to the reasons for the alternative claim. Conversely, the goal of deliberation can be elaborated through a set of subgoals: discuss each claim; present the possible arguments and counter-arguments for each claim; and form a conclusion (Nussbaum, 2008a).

Children have limited knowledge about the discourse elements that could comprise argument subgoals (Ferretti et al., 2009; Klein & Kirkpatrick, 2010). For example, elementary students frequently identify only one claim and a few reasons for it. They may show myside bias by presenting reasons for their own claim, omitting evidence against their claim, and omitting discussion of evidence for the alternative claim. These difficulties are particularly relevant to argumentation in science because students need to consider multiple conceptions and evaluate them on the basis of evidence. Prompting elementary writers with subgoals improves text quality and variety of argument moves (Ferretti, et al., 2009; Ferretti, MacArthur, & Dowdy, 2000; for a review, see Graham, McKeown, Kiuhara, & Harris, 2012). For example, Kuhn and Udell (2007) found that most elementary students did not spontaneously include counter-arguments in their writing, but, if prompted, they could generate them. Additionally, we suspected that students would not be sufficiently familiar with terms such as "persuade" or "discuss" to complete the task without further prompting. Thus, in the present study, to make the manipulation of argument goal types (persuasion versus deliberation) effective, it was necessary to provide subgoal prompts for various kinds of argument moves.

A question is how to present such subgoals. In one kind of distribution, the subgoals are presented in a single list. This type of prompt, which will be referred to as *clustered*, has been used in many persuasive writing studies (e.g., Ferretti et al., 2009; Nussbaum & Kardash, 2005). In a second kind of distribution, which we will call *segmented*, each writing subgoal prompt is presented separately, with space after each for the student to respond by outlining or drafting text (e.g., Harris, Graham, Mason, & Friedlander, 2007; Raison et al., 1994). Previous research in domains other than writing has shown that segmenting a learning activity into a series of separate subgoals allows students to attend to one subset of problem elements at a time, which reduces cognitive load and enhances learning (Ayres, 2006; Blayney, Kalyuga, & Sweller, 2010; Kester, Kirschner, & van Merriënboer, 2006). However, segmented and clustered subgoals have not been compared experimentally in the context of writing activities. Consequently, our second research question was whether a segmented series of subgoal prompts, compared to a clustered set, would differ in their effects on reasoning, cognitive load, and learning.

Interaction with Previous Knowledge: Expertise Reversals

Several previous studies concerning writing and "writing to learn" have found interactions between treatment conditions, and achievement or knowledge level (Gil et al., 2010; Nückles et al., 2010; Rivard, 2004). In cognitive load theory, interactions between cognitive load conditions and previous knowledge are referred to as *expertise reversals*: Students low in knowledge show lower cognitive load and greater learning in low cognitive load conditions, while students high in knowledge show a reversed pattern or a reduced response to the cognitive load treatment (e.g., Kalyuga & Renkl, 2010; Oksa, Kalyuga, & Chandler, 2010). Therefore, we posed the question of whether previous writing achievement would interact with type of writing goal and/or distribution of writing subgoals.

The Present Study

The study reported here focused on writing to learn about buoyancy. Buoyancy is a required topic in many elementary or secondary science curricula (e.g., Department for Education, United Kingdom, 2013; Ministry of Education for the Province of Ontario, 2007; University of the State of New York & State Education Department, n.d.; New Zealand Ministry of Education, n.d.). Students in late elementary school are expected to understand that the density of an object affects its buoyancy. Many children and adults hold the misconception that objects sink or float due to their weight alone, or they have an undifferentiated conception that confuses these two variables (Klein, 2000; She, 2005). Therefore, buoyancy is a topic that affords at least two explanations that would be plausible to many learners. Consequently, this topic presents a well-defined problem space that allows us to examine closely the way in which argument goals affect students' ability to identify claims, explore evidence, and draw valid inferences.

The design of the study was a pretest-post-test, between-subjects randomized experiment. First, students completed a pretest on buoyancy. Then, each student received a set of source documents representing possible claims and evidence. In a 2 x 2 x2 between-subjects design, participants low versus high in previous writing achievement were randomly assigned either a persuasion goal or a discussion goal, crossed with either a segmented or a clustered list of subgoal prompts. To recap, the research questions were the following:

1. What are the effects of writing goals (discussion versus persuasion), on reasoning in text, cognitive load, and buoyancy learning?
2. What are the effects of subgoal distribution (clustered versus segmented), on reasoning in text, cognitive load and buoyancy learning?
3. Do writing goals and/or distribution of subgoals interact with level of previous writing achievement to affect these variables?

Method

Participants and Curriculum Context

This study was conducted in a board of education serving a medium-sized city and the surrounding rural area. All students in four mixed Grade 7 and 8 classrooms and one Grade 8 classroom were invited to participate; 77 students responded to a letter of information by providing both student assent and parental consent and were included in the study. The sample included 30 Grade 7 and 47 Grade 8 students; they ranged in age from 12 years, 1 month to 14 years, 0 months. With respect to gender, 51 were female and 26 were male. One school was located in a rural area serving students of European heritage. The other four classes were located in two schools in urban neighborhoods serving lower and middle socioeconomic class students of European, Middle Eastern, and Asian heritage. Three of the students received English as a Second Language support at the time of the study. Four additional students had Individual Education Plans related to literacy.

For this study, students were categorized as either "low-achieving" or "high-achieving" with respect to writing based on their most recent end-of-term report grade. This was considered a valid estimate, because writing in this province is evaluated using rubrics and exemplars distributed by the provincial educational authority. Achievement level was defined by a median split: Students with report grades of 72% or less were defined as low-achieving; those with 73% or more were defined as high-achieving. This split approximated the provincial standard, which defined 70% as a satisfactory level of performance.

In classrooms with split (multiple) grades, a common practice in this board of education was for all students in a given class to complete the same science curriculum in a given year, with the entire class alternating between the Grade 7 curriculum one year and the Grade 8 curriculum the next year. In this year, the students were completing the Grade 8 science curriculum, which included a unit of study on Fluids. Each class completed the activity reported here at the beginning of the unit on Fluids, so that it would be relevant to the required curriculum, but would not be affected by possible variations in the teaching of the Fluids unit itself.

Materials

Pretest/post-test "Quiz on Floating and Sinking." The purpose of this instrument was to assess students' understanding of buoyancy. It comprised 10 items for a total of 16 points. Most items comprised two-tier multiple-choice items, in which students were presented with a novel scenario concerning buoyancy. They were asked to predict the outcome of the scenario, and select a reason to explain this prediction (Tsui & Treagust, 2010). Inter-item reliability was very good, $a = .73$; masked inter-rater reliability for scoring the short answer questions was 100.0 %. See Appendix A for a sample page from the post-test.

Buoyancy information package. This package provided information that students could use to infer the role of density in buoyancy, but did not present an explanation, so that students were required to reason about the evidence, form a claim, and create a unique argument text. The first page was titled "Weight, Volume and Density: They Are Not the Same." It presented a brief review of the concepts of weight and volume, and introduced the concept of density. The second section was titled "Objects that Sink or Float." It presented seven individual objects or pairs of objects: a cork; a large rock; the Emma Maersk (a large ship); a penny; two blocks of the same weight and different volumes; a kickboard; and two cylinders of the same volume but different weights. These were captioned with the following information: whether the object sank or floated in water; weight (grams or kilograms); volume (milliliters or liters); and a qualitative descriptor of the density of the object (dense/not dense). See Appendix B for a sample page from the information package. For the analysis below, it is important to note that some objects were critical in showing that density, and not weight alone, affects buoyancy (e.g., ship, penny, blocks); other objects were ambiguous, because either density, or weight alone, would support an accurate prediction about whether they floated or sank; they were often taken by students to support a weight explanation (e.g., cork, large rock, kickboard, cylinders).

Writing prompts. The purpose of the writing prompts was to implement the hypotheses with respect to the writing goal and distribution of writing subgoals see (Table 1). Each of the four sets of prompts combined a type of writing goal (discuss versus persuade) with a distribution of subgoals (segmented versus clustered).

Brief questionnaire on cognitive load. This questionnaire comprised two Likert scales concerning cognitive load, whose validity was supported by previous research (Brünken, Seufert, & Paas, 2010; DeLeeuw & Mayer, 2008). The first scale measured perceived difficulty: "How easy or difficult was this writing activity? Please circle a number," followed by a 9 point scale ranging from "Very very easy" to "very very difficult." The second Likert scale asked, "How much effort did you put into this writing activity? Please circle a number," followed by a 9 point scale that ranged from "very very little effort" to "very, very much effort."

Table 1

Writing Prompts, By Combination of Writing Goal and Subgoal Distribution

	Persuade Goal	Discuss Goal
Clustered Subgoals	Some people think that weight makes objects sink or float. Other people think that density makes objects sink or float. First, give your opinion. What makes objects float or sink? [2 lines] Next, persuade the reader that your opinion is right and the other opinion is wrong. Give as many reasons as you can. Use the objects to help you. [1 page]	Some people think that weight makes objects sink or float. Other people think that density makes objects sink or float. Discuss these two opinions. Use the objects to help you. Discuss as many as you can. [1 page] Finally, give your opinion: What makes objects float or sink? [2 lines]
Segmented Subgoals	Some people think that weight makes objects sink or float. Other people think that density makes objects sink or float. First, give your opinion: What makes objects float or sink? [2 lines] Persuade the reader of your opinion. Give as many reasons as you can. Use the objects to help you. [1/2 page] Write about an opinion different from yours. What reasons could someone give for this opinion? Give as many reasons as you can. Use the objects to help you. [1/4 page] Persuade the reader that the reasons for the other opinion are not good. Use the objects to help you. [1/4 page]	Some people think that weight makes objects sink or float. Discuss this opinion. Use the objects to help you. Discuss as many objects as you can. [1/2 page] Some people think that density makes objects sink or float. Discuss this opinion. Use the objects to help you. Discuss as many objects as you can. [1/2 page] Finally, give your opinion: What makes objects float or sink? [2 lines]

Procedure

All data collection was completed by a research assistant who was a graduate student in education and an elementary teacher, with the classroom teacher present.

Pretest. First, students completed the "Quiz on Floating and Sinking." All students did so in 15 minutes or less.

Writing activity. Five days later, students completed the writing activity. Each student was randomly assigned to one of four combinations of goal (persuade versus discuss) and distribution of subgoals (segmented versus clustered; see Table 1). Randomization was implemented by applying a random number table. Thus, each class included an approximately equal number of students who received each possible type of writing prompt. Students were allowed 25 minutes to complete the activity. All students completed it within this time. Immediately after writing, students completed the "Brief Questionnaire on Cognitive Load."

Post-test of buoyancy. Five days later, students completed the "Quiz on Floating and Sinking."

Textual Analysis

Texts were prepared by being word-processed to correct handwriting and spelling, both to conceal the treatment condition of each text and to minimize rater bias caused by the surface errors (Graham, Harris, & Hebert, 2011). Texts were then analyzed for bias/balance with respect to claim identification, evidence selection, and inference (see Tables 4, 5, and 6 for details). Claim identification represented the extent to which the student represented both her or his own claim plus an alternative claim, kappa = .75, $p < .001$. Evidence selection represented the extent to which the student presented a broad range of evidence, kappa = .81, $p < .001$. Inference balance represented the extent to which the writer validly supported one claim with adequate evidence as well as evaluating an alternative claim on the basis of evidence, kappa = .62, $p < .001$.

Results

Data Screening

The psychological dependent variables (perceived difficulty, effort, buoyancy learning) were normally distributed, so they were analyzed using a multivariate analysis of variance; the textual variables were nominal or ordinal, so they were analyzed using a series of chi-squared tests. Seventy-seven students consented to participate in the study. One student was withdrawn due to prolonged medical absence; three students scored at ceiling on the pretest; one student was an outlier with respect to gain score (> 3.0 SD). With these cases eliminated, the data from the remaining 72 students met all of the assumptions of MANOVA.

MANOVA of Psychological Variables

A MANOVA was used to test the effects argument goal type, subgoal distribution, and previous writing achievement upon the dependent variables of buoyancy learning, perceived difficulty, and effort. For the significant multivariate tests, univariate F-tests and Duncan's multiple range ad hoc tests were conducted as follow-up analyses (Table 3).

With respect to the first research question, the type of argument goal did not affect the combined dependent variable (see Table 2).

With respect to the second research question, subgoal distribution significantly affected the combined dependent variable (see Table 2). Follow-up ANOVAs indicated that it significantly affected perceived difficulty, F (1, 64) = 6.54, p =.01, partial $\eta2$ = .09, such that clustered subgoals were perceived to be more difficult, M = 4.93, SD = 1.65, than segmented subgoals, M = 4.11, SD = 1.75. Distribution of subgoals did not significantly affect buoyancy learning, F (1, 64) = .74, p =.39, partial $\eta2$ = .01; nor did it affect effort F (1, 64) = .86, p =.36, partial $\eta2$ = .01.

Previous writing achievement was not the focus of a research question, but it was an independent variable in the MANOVA. It significantly affected perceived difficulty, F (1, 64) = 23.26, p < .001, partial $\eta2$ = .27; the task was rated more difficult by low-achieving writers, M = 5.25, SD = 1.54, than high-achieving writers, M = 3.68, SD = 1.57. Previous writing achievement

Table 2

MANOVA of the Effects of Argument Goal Type, Subgoal Distribution, and Previous Writing Achievement on the Clustered Dependent Variable (Buoyancy Learning, Effort, Perceived Difficulty)

Effect	Wilk's Lambda	*F*	*df*	*p*	Partial eta squared
Writing achievement	.699	8.90	3	<.001	.30
Writing goal type	.926	1.65	3	.186	.07
Subgoal distribution	.881	2.80	3	.047*	.12
Writing achievement x Writing goal	.977	.50	3	.686	.02
Writing achievement x Subgoal distribution	.925	.17	3	.198	.08
Writing goal x Subgoal distribution	.979	.45	3	.717	.02
Writing achievement x Writing goal x Subgoal distribution	.841	3.91	3	.013*	.16
Error			62		

Table 3

Effort, Difficulty, and Buoyancy Learning By Previous Writing Achievement, Writing Goal Type, and Subgoal Distribution

	Low achieving writers				High-achieving writers			
	Persuade goal		Discuss Goal		Persuade Goal		Discuss Goal	
	Cluster subgoals	Segment subgoals	Cluster subgoals	Segment subgoals	Cluster subgoals	Segment subgoals	Cluster subgoals	Segment subgoals
Effort	6.70^{ab} (2.16)	5.60^{a} (1.84)	5.38^{a} (1.77)	5.70^{a} (1.34)	5.88^{a} (1.55)	7.50^{b} (1.07)	6.33^{ab} (1.12)	6.89^{ab} (1.54)
Difficulty	5.35^{c} (1.60)	4.30^{bc} (1.83)	5.75^{c} (1.39)	5.70^{c} (.95)	3.40^{ab} (1.51)	3.50^{ab} (1.20)	5.11^{c} (1.27)	2.67^{a} (1.32)
Buoyancy learning	2.00 (2.26)	2.60 (2.76)	2.50 (2.39)	1.70 (1.77)	4.75 (3.01)	2.88 (2.36)	2.78 (2.22)	2.89 (2.47)

Note: Means not sharing a common superscript differ significantly from one another, p <.05

also affected buoyancy learning, F (1, 64) = 3.86, p = .05, partial $\eta 2$ = .06, such that high-achieving writers made greater gains, M = 3.29, SD = 2.54, than low-achieving writers, M = 2.18, SD = 2.25. Previous writing achievement also significantly affected effort, F (1, 64) = 4.51, p = .04, partial $\eta 2$ = .07, such that high achieving writers reported higher effort, M = 6.65, SD = 1.41, than low-achieving writers, M = 5.87, SD = 1.80.

The third research question concerned interactions between previous writing achievement and treatment conditions. The two-way interactions were not statistically significant (see Table 2). The three-way interaction of goal type by subgoal distribution by previous writing achievement significantly affected perceived difficulty, F (1, 64) = 7.13, p =.01, partial $\eta 2$ = .10 (see Table 3). High-achieving writers rated the discuss goal + clustered subgoals condition significantly more difficult than the discuss goal + separate subgoals condition; low achieving writers rated the all of the conditions to be similarly difficult. The three-way interaction also marginally affected effort, F (1, 64) = 2.70, p =.10, partial $\eta 2$ = .04 (see Table 3); post-hoc comparisons showed that the high-achieving writers in the persuade goal + segmented subgoals condition reported significantly more effort than high achieving writers in the persuade goal + clustered subgoals condition; low achieving writers rated their effort similarly across all conditions (see row 2). The three way interaction did not significantly affect science learning, F (1, 64) = 2.19, p =.14, partial $\eta 2$ = .03.

Textual Analysis of Reasoning

Recall that three aspects of the text were analyzed for bias versus balance: claim identification, evidence selection, and inference. See Appendix C for samples of four texts, categorized with respect to bias/balance in claim identification, evidence selection, and inference.

Claim identification. The type of argument goal significantly affected claim identification, $\chi 2$ (2, N = 72) = 9.33, p = .009 (See Table 4), such that more students in the discussion condition, compared to the persuasion condition, made no clear claim. More students in the persuasion condition than discussion condition showed myside claim bias by asserting one claim and identifying no alternative claim. Approximately equal proportions of students in the two types of goal conditions showed balance by asserting one claim and identifying an alternative claim.

Distribution of subgoals had a large effect on claim identification χ^2 (2, N = 72) = 11.03, p = .002 (see Table 4). More students in the segmented subgoal condition, compared to the clustered subgoal condition, presented no clear claim; more writers in the clustered subgoal condition showed myside claim bias by discussing only one claim; and more writers in the segmented condition showed balanced claim identification by stating a claim and identifying an alternative claim.

Previous writing achievement marginally affected claim identification, $\chi 2$ (2, N = 72) = 3.99, p = .07, with a significant linear-by-linear association, $\chi 2$ (1, N = 72) = 3.83, p = .03 (see Table 4). Low achieving writers were more likely than high-achieving writers to present no clear claim; high-achieving writers were somewhat more likely to show balanced claim identification.

Table 4

Claim Identification, By Goal Type, Subgoal Distribution, and Previous Writing Achievement.

Claim Identification	Goal Type (N = 72)		Subgoal Distribution (N = 72)		Previous Writing Achievement (N = 72)	
	Persuasion (n = 36)	Discussion (n = 36)	Cluster (n = 35)	Segment (n = 37)	Low (n = 38)	High (n = 34)
No clear claim. Text includes no clear thesis; e.g., missing, multiple, or contradictory claims.	2	11	6	7	10	3
Myside claim bias. Text asserts one claim, identifies no alternative claim.	7	2	9	0	5	4
Claim balance. Text asserts writer’s claim plus identifies an alternative claim.	27	23	20	30	23	27

Table 5

Bias/Balance in Evidence Selection, By Goal Type, Subgoal Distribution, and Previous Writing Achievement

Level of Evidence Selection	Goal Type (N = 72)		Subgoal Distribution (N = 72)		Previous Writing Achievement (N = 72)	
	Persuasion (n = 36)	Discussion (n = 36)	Cluster (n = 35)	Segment (n = 37)	Low (n = 38)	High (n = 34)
No evidence. Evidence to support writer's claim is lacking.	6	3	5	4	8	1
Myside Bias. Evidence has ambiguous implications, (e.g., large rock, cork) but is presented to support writer's claim.	4	9	7	6	8	5
Valid selection. Evidence sufficient to distinguish between possible claims; e.g., blocks, large ship, penny.	5	5	6	4	5	5
Balanced Evidence. Text includes critical evidence, as well as ambiguous evidence that could be taken to support the alternative claim.	21	19	17	23	17	23

Evidence selection. The type of writing goal did not significantly affect evidence selection, χ^2 (3, N = 72) = 3.02, p = .39 (see Table 5). The distribution of subgoals also did not significantly affect evidence selection, χ^2 (3, N = 72) = 1.43, p = .70 (Table 5).

Previous writing achievement significantly affected evidence selection, χ^2 (3, N = 72) = 6.84, p = .04, such that low achieving writers were more likely than high achieving writers to present no evidence, while high-achieving writers were more likely to show balance by discussing a wide range of evidence.

Inference balance. Goal type significantly affected inference bias/balance, χ^2 (3, N = 72) = 8.69, p = .03 (see Table 6). More students in the discussion condition than the persuasion condition made no clear claim, so that their inference bias could not be evaluated; more students in the persuasion condition showed myside inference bias by asserting claims based on inadequate or contrary evidence. More students in the persuasion condition also showed unbiased inference by validly basing their claim on critical evidence. A similar proportion of writers in the persuasion and discussion goal conditions showed inference balance by validly discussing the evidence for two different claims (Table 6).

Distribution of subgoals marginally affected inference bias, χ^2 (3, N = 72) = 5.15, p = .08 (see Table 6). The clustered subgoal condition, relative to the segmented subgoal condition, resulted in somewhat more texts with unbiased valid inferences concerning one claim; the segmented subgoal condition resulted in somewhat more texts with balanced inferences that validly related two different claims to evidence.

Table 6

Bias/Balance in Inference, By Goal Type, Subgoal Distribution, and Previous Writing Achievement

Inference Bias/Balance	Goal Type (N = 72)		Subgoal Distribution (N = 72)		Previous Writing Achievement (N = 72)	
	Persuasion (n = 36)	Discussion (n = 36)	Cluster (n = 35)	Segment (n = 37)	Low (n = 38)	High (n = 34)
No clear claim. There may be multiple or contradictory claims with no resolution.	2	11	6	7	10	3
Myside inference bias. Claim is supported by contrary evidence, no evidence, or insufficient evidence.	11	6	8	9	11	6
Unbiased (valid) inference. Claim is supported by logically sufficient evidence, which distinguishes it from possible alternative claims.	16	11	17	10	14	13
Balanced inference. Meets criteria for "Unbiased inference," plus relates alternative claim to evidence.	7	8	4	11	3	12

Previous writing achievement strongly affected inference balance, χ^2 (3, N = 72) = 10.48, p = .01 (see Table 6). More low-achieving writers, compared to high-achieving writers, produced texts with no clear claims. More low-achieving writers also produced texts with myside inference bias. Approximately equal numbers of low and high achieving writers produced unbiased texts with valid evidence for one claim. More high achieving writers than low achieving writers produced texts with balanced inferences that supported one claim with evidence, while also validly discussing the evidence with respect to an alternative claim (Table 6).

Discussion

Argumentation is a means of thinking critically and learning in science. This study was motivated by previous research, which showed that argumentation goals affect the variety of argument moves that students make in written text, and that argument goals affect learning through verbal discussion. This study was the first to investigate: the effect of two types of argument goals on individual writing, reasoning, and learning in science; the effect of segmented versus clustered subgoals; as well as the interaction of these two variables with previous writing achievement.

Effects of Argument Goal Type

The first question concerned the effect of argument goals on reasoning, cognitive load, and learning. The results showed that discussion goals had a double-edged effect on reasoning. As expected, discussion goals reduced instances of myside bias in both claims and inferences. This is broadly consistent with previous research, which has shown that discussion goals tend to elicit arguments that are relatively more dialectical than persuasive goals. However, a new finding was that discussion goals also increased the number of texts with no coherent claim, as some writers presented conflicting explanations without resolving them. These incoherent texts came primarily from low-achieving writers. This is consistent with the fact that argument-counterargument integration is challenging for students (Nussbaum, 2008b; Nussbaum & Schraw, 2007). It appears that once discussion goals elicit discussion of competing claims, lower-achieving writers have difficulty formulating a text that integrates discussion of these claims into one coherent text. Additionally, persuade and discuss goals elicited equal numbers of texts that presented balanced claims and balanced inferences, in which writers validly related at least two different claims to evidence. As qualifications to this finding, we note that balanced texts more often came from high-achieving students.

Conversely, argument goals did not affect learning. One might expect that more balanced discourse would have resulted in increased learning. However, it appears that for most students the prompt to argue on the basis of evidence was sufficient to elicit learning, whether the overall goal was framed in terms of persuasion or discussion. In their written texts and on the post-test, most students adopted a density conception, rather than holding to their original claim. Thus it appears that the persuasion/discussion distinction may impact students' written expression more strongly than it impacts their actual understanding of subject matter.

Effects of Subgoal Distribution

The second question concerned the effects of subgoal distribution on reasoning, cognitive load, and learning. The theoretical rationale for segmenting subgoals was based on cognitive load research in domains other than writing (Ayres, 2006; Blayney, Kalyuga, & Sweller, 2010). In the present study, with respect to the reasoning variables, the segmented subgoal condition significantly affected claims and inferences. Consistent with our rationale, the present study showed that subgoal segmentation reduced cognitive load as measured by perceived difficulty. Thus, the present study contributes to the nascent literature showing that cognitive load effects may be found not only in well-defined task domains such as mathematics, but also in less well-defined domains such as English studies and written composition (cf., Oksa et al., 2010).

Segmented subgoals resulted in fewer texts with myside claim bias, and more texts with balanced claim identification. They also resulted in marginally more texts with balanced inferences, that is, evidence-based discussions of at least two claims. This supports the use of segmented subgoal presentation for introducing students to difficult writing tasks (this was particularly important for the discussion goal; see below). However, segmentation also produced some incoherent texts, particularly for low achieving writers. This again suggests that once we elicit a consideration of conflicting opinions from low-achieving writers, they may have difficulty resolving these opinions to form a coherent text.

However, there was no effect of subgoal distribution on learning. This may be attributed to the fact that this segmentation elicited more texts that were balanced with respect to inferencing, which might be expected to contribute to learning; however, segmentation also elicited more texts that were incoherent, which might be expected to detract from learning. Possibly these trends cancelled one another out.

Expertise Reversal Interactions

Our third research question concerned interactions between previous writing achievement and writing goals, as well as between previous writing achievement and subgoal distribution. With respect to perceived difficulty, the high-achieving writers rated the "discuss goal + clustered subgoals" more difficult than the "discuss goal + segmented subgoals" condition. We interpret this in light of the fact that the discussion goal was rated higher in difficulty than the persuasion goal, perhaps because it appears less often in curriculum documents. It appears that when addressing this novel genre, high-achieving writers benefited from being supported step by step with segmented subgoals. Conversely, low achieving writers tended to rate all writing conditions relatively difficult.

Effort showed a complementary, marginally significant trend. For high-achieving writers, with persuasive text, segmented goals required *more* effort than clustered subgoals. This is consistent with cognitive load theory, in which expertise reversal is interpreted as a redundancy effect: For knowledgeable students, schemata in long-term memory provide procedural guidance and content information for a given task. Scaffolding can disrupt this process, requiring students to use working memory to reconcile the prescribed procedure with their existing knowledge (Kalyuga & Renkl, 2010; Sweller et al., 2011). In the case of persuasive argument writing, skilled upper elementary writers would have some discourse knowledge about this text genre in long-term memory (e.g., Klein & Kirkpatrick, 2010). The segmented subgoal condition, which imposed a specific sequence of writing subgoals, may have disrupted this and required the learners to integrate the prescribed sequence with their prior genre knowledge. This was experienced as effortful. Conversely, the clustered subgoal condition does not present the same pressure for writers to follow a given sequence. The present study extends the small literature on expertise reversals in writing (Gil et al., 2010; Nückles et al., 2010).

Limitations

The main limitation of this study is that the data set, while large enough for detecting main effects, was relatively small for detecting interactions. The number of participants per cell for the three-way interaction was small, so the error estimates are large. This means that caution must be used in generalizing from these results. A second limitation of the present study is that it was based on a well-defined question in the in the domain of physics. It would be desirable to further investigate the effects of goal type on learning about a controversial social issue (e.g., Felton et al., 2009; Nussbaum, 2005; Wolfe et al., 2009). A third limitation is that we did not collect data on the students' previous writing experiences in science; however, we note that writing in science is required in this province (Ministry of Education for the Province of Ontario, 2007).

A fourth limitation arises from the relative difficulty of discussion as an argument goal. In this study, this unfamiliarity was mitigated using subgoal prompts; these were effective for high-achieving writers in the segmented subgoal condition, resulting in "easy" ratings for the segmented version of the task. However, lower achieving writers found all writing conditions, including segmented subgoal conditions, more difficult. A different approach, which may support low-achieving writers more strongly, would be to teach students a strategy for writing a discussion, and to compare the effect of this kind of argument with persuasive argumentation. De La Paz and Graham (1997) have created a writing strategy of this kind called STOP and

DARE, which has been used effectively for writing to learn in history (De La Paz & Felton, 2010).

Educational Implications

Pending replication, these results point to some possible educational implications:

- Teachers could consider assigning students to write arguments about competing explanations, in order to elicit critical thinking and conceptual change in science. Recently, researchers have developed programs that integrate argument writing with related activities such as discussion, reading, and planning experiments with positive results (Hand et al., 2004; McNeill et al., 2009). Additionally, computer supported collaborative learning platforms have incorporated prompts for dialogical argumentation (e.g., Yeh & She, 2010).
- Students could be assigned to address discussion goals rather than persuasion goals to reduce myside bias in claims and inferences. This can be expected to elicit a wider variety of argumentation moves (Felton et al., 2013; Keefer, Zeitz & Resnick, 2000; Nussbaum, 2005; Nussbaum & Kardash, 2005). However, our results indicate that lower-achieving writers may have difficulty integrating arguments and counter-arguments to form a coherent text. An alternative approach to individual writing of dialogical arguments can be to elicit oral discussion prior to writing (Felton et al., 2009).
- Students could be provided with subgoals in a segmented format in order to reduce cognitive load and reduce myside bias in claims and inferences. Although this was the first direct experimental comparison of segmented versus clustered subgoals, several previous studies and resources have employed segmented goals (Harris et al., 2007; Raison et al., 1994). Evidence-based resources are available that provide strategies comprised of subgoals for argumentation, as well as other nonfiction genre such as explanation and comparison (Harris et al., 2007).
- For higher-achieving (above average) students writing in a familiar genre (e.g., persuasive argumentation), teachers could provide subgoal prompts in a clustered format rather than a segmented format to avoid expertise reversals; for less familiar genres, such as discussion argumentation, they could present subgoals in a segmented format.

Conclusion

This was the first study to investigate the effects of persuasion and discussion goals on learning during individual writing. The results largely supported the view that discussion goals reduce myside bias in claim identification and inference. However, this was complicated by the fact that discussion goals also produced more texts with incoherent claims and inferences. Subgoal segmentation affected reasoning and reduced cognitive load. Goal type interacted with previous achievement and subgoal segmentation to affect cognitive load, such that it appears that for high achieving writers, for a less familiar genre, segmented subgoals reduced cognitive load, but for a familiar genre of persuasion, they increased it. Theoretically, we can understand these findings by considering that goals affect writing through the subgoals selected by the writer. In this sense, science argumentation could be conceived as a process that results from students' interpretation of goals, elaborated by subgoals, according to their level of writing achievement.

Acknowledgements

The researchers wish to thank the research officers, administrators, teachers, and students of the Thames Valley District School Board and the London Catholic District School Board for their help with this project.

This research was supported by the Social Science and Humanities Research Council of Canada, Insight Grant 435-2012-0687.

References

Asterhan, C. S. C., & Schwarz, B. B. (2007). Effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology, 99*, 626-639. doi: 10.1037/0022-0663.99.3.626

Ayres, P. (2006). Impact of reducing cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology, 20*, 287-298. doi: 10.1002/acp.1245

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29-58.

Blayney, P., Kalyuga, S., & Sweller, J. (2010). Interactions between the isolated-interactive elements effect and levels of learner expertise: Experimental evidence from an accountancy class. *Instructional Science, 38*, 277-287. doi:10.1007/s11251-009-9105-x

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. L . Plaas, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 181-202). Cambridge, United Kingdom: Cambridge University Press.

Cavagnetto, A. R. (2010). Argument to foster scientific literacy: A review of argument interventions in K-12 science contexts. *Review of Educational Research, 80*, 336-371.

De La Paz, S., & Felton, M. K. (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology, 35*, 174-192. doi:10.1016/j.cedpsych.2010.03.001

De La Paz, S., & Graham, S. (1997). Effects of dictation and advanced planning instruction on the composing of students with writing and learning problems. *Journal of Educational Psychology, 89*, 203.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*, 223-234. doi:10.1037/0022-0663.100.1.223

Department for Education. (2013). *Science programmes of study: Key stage 3: National curriculum in England*. United Kingdom: The National Archives. Retrieved from http://dera.ioe.ac.uk/18294/1/SECONDARY_national_curriculum_-_Science.pdf

Duke, N. K., Caughlan, S., Juzwik, M. M., & Martin, N. M. (2012). *Reading and writing genre with purpose in K-8 classrooms*. Portsmouth, NH: Heinemann.

Felton, M., Garcia-Mila, M., & Gilabert, S. (2009). Deliberation versus dispute: The impact of argumentative discourse goals on learning and reasoning in the science classroom. *Informal Logic, 29*, 417-446.

Ferretti, R. P., Lewis, W. E., & Andrews-Wecherly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology, 101*, 577-589. doi:10.1037/a0014702

Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology, 92*, 694-702. doi:10.10371/13022-0663.92.4.694

Garcia-Mila, M, Gilabert, S., Erduran, S., & Felton, M. (2013). The effect of argumentative task goal on the quality of argumentative discourse. *Science Education, 97*, 497-523. doi:10.1002/sce.21057

Gil, L., Bråten, I. Vidal-Abarca , E., & Strømsø, H. I. (2010). Summary versus argument tasks when working from multiple documents: Which is better for whom? *Contemporary Educational Psychology, 35*, 157-173. doi:10.1016/j.cedpsych.2009.11.002

Gillespie, A., Olinghouse, N. G., & Graham, S. (2013). Fifth-grade students' knowledge about writing process and writing genres. *The Elementary School Journal, 113*, 565-588. doi: 10.1086/669938

Golanics J. D. & Nussbaum, E. M. (2008). Enhancing online collaboration through question elaboration and goal instructions. *Journal of Computer Assisted Learning, 24*, 167-180.

Graham, S., Harris, K. R., & Hebert, M. (2011). It is more than just the message: Presentation effects in scoring writing. *Focus on Exceptional Children, 44*, 1-13

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*, 879-896. doi:10.1037/a0029185

Hand, B., Wallace, C. W., & Yang, E. M. (2004). Using a science writing heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: Quantitative and qualitative aspects. *International Journal of Science Education, 26*, 131-149. doi:10.1080/0950069032000070252

Harris, K. H, Graham, S., Mason, L. H., & Friedlander, B. (2007). *Powerful writing strategies for all students*. Baltimore, MD: Brookes Publishing.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*, 369-388. doi: 10.1177/0741088312451260

Hayes, J. R., & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication, 23*, 135-149. doi: 10.1177/0741088306286283

Hebert, M., Gillespie, A., & Graham, S. (2013). Comparing effects of different writing activities on reading comprehension: A meta-analysis. *Reading and Writing, 26*, 111-138. doi: 10.1007/s11145-012-9386-3

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychology, 38*, 23–31.

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science, 38*, 209-215. doi:10.1007/s11251-009-9102-0

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

Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2006). Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology, 31*, 167-185. doi:10.1016/j.cedpsych.2005.04.002

Kieft, M., Rijlaarsdam, G., & van den Bergh, H. (2008). An aptitude-treatment interaction approach to writing-to-learn. *Learning and Instruction, 18*, 379-390. doi:10.1016/j.learninstruc.2007.07.004

Klein, P. D. (2000). Elementary students' strategies for writing-to-learn in science. *Cognition and Instruction, 18*, 317-348.

Klein, P. D., & Kirkpatrick, L. C. (2010). A framework for content area writing: Mediators and moderators. *Journal of Writing Research, 2*(1), 1-46.

Kroll, B.M. (2005). Arguing differently. *Pedagogy, 5*(1), 37-60.

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

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

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child development, 74*, 1245-1260. doi:10.1002/sce.20395

Langer, J. A., & Applebee, A. N. (1987). *How writing shapes thinking: A study of teaching and learning*. Urbana, IL: National Council of Teachers of English.

Lewis, W. E., & Ferretti, R. P. (2011). Topoi and literary interpretation: The effects of a critical reading and writing intervention on high school students' analytic literary essays. *Contemporary Educational Psychology, 36*, 334-354. doi:10.1016/j.cedpsych.2011.06.001

McCutchen, D., Teske, P., & Bankston, C. (2008). Writing and cognition: Implications of the cognitive architecture for learning to write and writing to learn. In C. Bazerman (Ed.) *Handbook of research on Writing: History, society, school, individual, text* (pp. 451-465). New York, NY: Lawrence Erlbaum Associates.

McNeill, K. L., & Krajcik, J. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain-specific and domain-general knowledge in writing arguments to explain phenomena. *The Journal of the Learning Sciences, 18*, 416-460. doi:10.1080/10508400903013488

Mercer, N. (2000). *Words and minds: How we use language to think together*. London, United Kingdom: Routledge.

Ministry of Education. (2006). *The Ontario Curriculum, Grades 1-8: Language (revised)*. Toronto, ON: Queen's Printer for Ontario.

Ministry of Education. (2007). *The Ontario Curriculum, Grades 1-8: Science and technology*. Toronto, ON: Queen's Printer for Ontario.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington, DC: National Governors Association. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

New Zealand Ministry of Education (n.d.). *Ti kete ipurangi. Book 38: Understanding Buoyancy: Why objects sink or float*. Wellington, New Zealand: New Zealand Ministry of Education.

Nückles, M., Hübner, S., Dümer, S., & Renkl, A. (2010). Expertise reversal effects in writing-to-learn. *Instructional Science, 38*, 237-258. doi:10.1007/s11251-009-9106-9

Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction, 19*, 259-271. doi:10.1016/j.learninstruc.2008.05.002

Nussbaum, E. M. (2005). The effect of goal instructions and need for cognition on interactive argumentation. *Contemporary Educational Psychology, 30*, 286-313. doi:10.1016/j.cedpsych.2004.11.002

Nussbaum, E. M. (2008a). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology, 33*, 345-359. doi:10.1016/j.cedpsych.2008.06.001

Nussbaum, E. M. (2008b). Using argumentation vee diagrams (avds) for promoting argument-counterargument integration in reflective writing. *Journal of Educational Psychology, 100*, 549-565. doi:10.1037/0022-0663.100.3.549

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

Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *The Journal of Experimental Education, 76*, 59-92.

Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology, 28*, 384-395. doi:10.1016/S0361-476X(02)00038-3

Nussbaum, E. M., Sinatra, G. M., & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education, 30*, 1977-1999. doi:10.1080/09500690701545919

Oksa, A., Kalyuga, S., & Chandler, P. (2010). Expertise reversal effect in using explanatory notes for readers of Shakespearean text. *Instructional Science, 38*, 217-236. doi:10.1007/s11251-009-9109-6

Olive, T., Kellogg, R. T., & Piolat, A. (2008). Verbal, visual, and spatial working memory demands during text composition. *Applied Psycholinguistics, 29*, 669-687. doi:10.1017/S0142716408080284

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*, 463-466. doi:10.1126/science.1183944

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science.

Journal of Research in Science Teaching, 41, 994-1020. doi:10.1002/tea.20035
Raison, G., Rivalland, J., Derewianka, B., Johnson, T. D., Sloan, P., Latham, R., Kovalevs, K., Bindon, R., Annandale, K., & Dewsbury, A. (1994). *First steps: Writing resource*. Portsmouth, NH: Heinemann.
Rapanta, C., Garcia-Mila, M., & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research, 83*, 483-520. doi:10.3102/0034654313487606
Rivard, L.P. (2004). Are language-based activities in science effective for all students, including low achievers? *Science Education, 88*, 420-442. doi:10.1002/sce.10114
Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education, 88*, 345-372. doi:10.1002/sce.10130
Schwarz, B. B. (2009). Argumentation and learning. In N. M. Mirz & A-N. Perret-Clermont (Eds.), *Argumentation and education: Theoretical foundations and practices* (pp. 91-126). New York, NY: Springer.
She, H.C. (2005). Enhancing eight grade students' learning of buoyancy: The interaction of teacher' instructional approach and students' learning preference style. *International Journal of Science and Mathematics Education, 3*, 609-624.
Shehab, H. M., (2011). *Cognitive load of critical thinking strategies* (Doctoral dissertation). Retrieved from http://digitalscholarship.unlv.edu/thesesdissertations/943/
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*, 257-285. doi:10.1016/0364-0213(88)90023-7
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.
The University of the State of New York and the State Education Department. (n.d.). *Intermediate level science core curriculum, Grades 5-8*. Albany, NY: New York State Education Department.
Tsui C.-Y. & Treagust, D. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education, 32*, 1073-1098. doi:10.1080/09500690902951429
Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16-26. doi:10.1080/00461520701756248
Wiley J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*, 301-311. doi:10.1037/0022-0663.91.2.301
Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication, 26*, 183-209. doi:10.1177/0741088309333019
Yeh, K. H., & She, H. C. (2010). On-line synchronous scientific argumentation learning: Nurturing students' argumentation ability and conceptual change in science context. *Computers & Education, 55*, 586-602. doi:10.1016/j.compedu.2010.02.020
Zhu, X., & Zhang, J. (2005). The influence of "goal-free effect" and "worked examples" on students' writing achievement. *Psychological Science (China), 28*, 1139-1143. (extended abstract trans. by X. Zhu in personal communication).
Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*, 35-62. doi:10.1002/tea.10008

Dr. Perry Klein is an Associate Professor in the Faculty of Education at Western University, where he currently serves as Chair of Applied Psychology. He teaches courses in educational psychology and literacy. His research focuses on the psychology of writing and writing instruction.

Jacqueline Ehrhardt is a Primary School Teacher in London, England. She completed her Master of Education degree at Western University where she focused her thesis research on revision within the primary classroom.

Appendix A: Sample Page from Post-test

5. Eric weighs 50 kilograms and he is very muscular. Fred weighs 70 kilograms and he is less muscular, with more fat. Fat is less dense than muscle.
Who is more likely to float?

    (a) Eric
    (b) Fred
    c) they will both float

because…

    a) Fred is less dense
    b) Eric weighs less
    c) everyone floats

6. A submarine is under the water. It is not rising or sinking. It has tanks full of water. Then, air is used to blow the water out of the tanks. Now the tanks are full of air. What will it do?

    a) rise
    b) sink
    c) neither rise nor sink

because…

    a) the submarine now weighs less but has the same volume as before
    b) the submarine now has air in it
    c) it is the same submarine

7. There is a mystery liquid. It begins to cool and turn solid. When it becomes solid, it expands (grows larger). What will the solid part do?

    a) float on the liquid part
    b) sink in the liquid part
    c) neither float nor sink

because…

    a) it is more dense now
    b) it is less dense now
    c) it is the same material as before

Appendix B: Sample Page from Buoyancy Information Package

Objects that Float or Sink

(Please return this)

This cork floats.

Weight: 2 grams
Volume: 9 cubic centimeters

It is not dense.

This rock sinks.

Weight: 300 kilograms
Volume: 100 liters

It is dense.

The Emma Maersk is one of the world's largest ships. It floats.

Weight: 156, 907, 000 kilograms
Volume: 170, 000, 000 liters

Overall, it is not dense.

This penny sinks.

Weight: 2.35 grams
Volume: 0.30 cubic centimeters

It is dense.

Appendix C: Samples of Arguments, Categorized on Three Dimensions of Balance/Bias

Amanda (#12) Claim balanced; evidence balanced; inference balanced.

I think that both density and weight makes something float or sink, because density means how much something weighs for its size. …For example, the Emma Maersk is 170,000,000 L (very big), but it weighs 156, 907,000. For its size, it weighs very little, so it floats. A penny, however, if it is only 30 cubic cm, but 2.35 grams, it will sink because it is heavy for its size.

Someone might say that just weight makes something float or sink because when things are heavy they sink and when they are light they float. For instance, a cork floats because it is light, and a rock sinks because it is heavy.

What about the size though? The Emma Maersk is very heavy, but it floats and the penny is very light, but it sinks. That is why the other opinion is not good.

Jed (#27) Claims balanced; evidence biased; inference biased

An object that is dense, it sinks. An object that is not dense it floats.

If you put a rock in water it will sink because it weighs 300 kilograms and its volume is 100 liters. When you put a cork in water, it will float because its weighs 2 grams and its volume is 9 cubic cm.

Most light things float because they are not dense and most heavy things sink because they are denser.

They [the other opinion] is not good because they can be wrong.

Stacey (#25) Claim balanced; no evidence; inference biased.

I think weight makes stuff sink and float because if something is heavy it will sink and if it is light it will float.

If something is heavy it will sink and if it is light it will float.

If something is heavy it will sink because it puts more pressure on the water witch results in forcing the water to give up. If something is light it will float because it puts less pressure on the water so it wont force the water to give up.

Density makes things sink because its heavy and large, it puts pressure to force the water to give up.

Their (other) opinion is not good because they don't have any back-up.

Michael (#17) Claim biased; evidence unbiased; inference unbiased.

I say that the answer is density, because density deals with weight and volume, density is how heavy an object is for its size.

My opinion is right because density deals with the volume and weight. If something weighs more than its volume of water, it floats. If it weighs less than its volume of water is sinks.

An example of this is a ship, lets's say the Emma Maersk. The volume of the ship is 170000000liters and the weight of the ship is 156000000. Because the ship weighs less than the volume of water it floats.

[Presents penny as further evidence]
[No mention of alternative claim]