2015

Effects of persuasion and discussion goals on writing, cognitive load, and learning in science.

Perry Klein
Western University, pklein@uwo.ca

J. S. Eharhardt

Citation of this paper:
Klein, P. D., & Ehrhardt, J. S. (2015). Effects of Persuasion and Discussion Goals On Writing, Cognitive Load, and Learning in Science. Alberta Journal of Educational Research, 61(1), 40-64.

THE EFFECTS OF RHETORICAL AND CONTENT SUBGOALS ON WRITING AND LEARNING

Publication reference: Klein, P. D., & Ehrhardt, J. S. (2015). Effects of persuasion and discussion goals on writing, cognitive load, and learning in science. *Alberta Journal of Educational Research*. 61(1), 40-64.

Abstract

Argument writing is challenging for elementary students. Previous experimental research has focused on scaffolding rhetorical goals, leaving content goals relatively unexplored. In a randomized experiment, 73 students in Grades 5, 6, and 7 wrote persuasive texts about difficult-to-classify vertebrates. Each student received one of three sets of writing prompts: a persuasive goal only (control); persuasive goal + rhetorical subgoal prompts; or persuasive goal + content subgoal prompts. Rhetorical subgoals increased text quality, variety of rhetorical moves, number of complex propositions, and classification knowledge. Content subgoals increased the number of simple propositions in text. A path analysis indicated that content subgoal prompts and rhetorical subgoal prompts elicited different paths to writing and learning.

Keywords: Writing; writing goals; writing to learn; cognitive processes; argumentation; science learning.

Introduction

The Problem

Argument writing can be an effective means of communicating and learning in the subject areas (Wiley & Voss, 1999). However, elementary students find argument writing challenging (Ferretti, Lewis & Andrews-Weckerly, 2009; Means & Voss, 1996). The difficulty of argument writing for many students affects both their text quality, and their ability to use argumentation as a means of learning (Klein & Rose, 2010; Klein & Kirkpatrick, 2010). Several studies have shown that scaffolding elementary students by prompting rhetorical goals can contribute to the holistic quality of their texts (Ferretti et al., 2009; Graham, McKeown, Kiuhara, & Harris, 2012); however, to date, experimental research on the effects of prompting rhetorical goals on learning has been limited (Hebert, Gillespie & Graham, 2013). Additionally, initial research suggests that prompting content goals may contribute to writing and learning (Butcher & Kintsch, 2001; McNeill & Krajcik, 2009); however, nearly all experimental research on elementary writing has focused on prompting rhetorical goals rather than content goals (see Graham et al., 2012). The present study investigated the effects of rhetorical and content goal prompting on writing, cognitive load, and learning in elementary science.

Argumentation in Content Area Writing and Learning

In the content areas, argumentation can allow students to communicate their ideas, to think critically about controversial topics, and to improve their understanding of these topics (Cavagnetto, 2010; Schwarz, 2009; Wiley & Voss, 1999). In science for example, argumentation is an important part of professional discourse. It has also become an important part of science education (Cavagnetto, 2010; Kuhn, 2010; Osborne, 2010). Argumentation can allow students to consider various conceptions, and choose among them on the basis of evidence (Asterhan &

Schwarz, 2007; Nussbaum & Sinatra, 2003; Nussbaum, Sinatra & Poliquin, 2008). To date, most studies of argumentation in science have focused on face-to-face or online dialogue (e.g., Asterhan & Schwarz, 2007; Chen & She, 2012; see Cavagnetto, 2010 for a review). However, writing can also be an effective modality for learning (Bangert-Drowns, Hurley & Wilkinson, 2004; Hebert et al., 2013). Argument writing has been incorporated into several science education methods (e.g., Bell & Linn, 2000; Chen & She, 2012; Hand, Wallace & Yang, 2004).

However, the difficulty of argument writing for elementary and secondary students affects their ability to use argumentation as a means of learning (Gil, Bråten, Vidal-Abarca & Strømsø, 2010; Klein & Kirkpatrick, 2010; Means & Voss, 1996). Argument writing includes two dimensions, which will be referred to here as *rhetoric* and *content* (Bereiter & Scardamalia, 1987; Butcher & Kintsch, 2001; for related distinctions, see Means & Voss, 1996; Schworm & Renkl, 2007). For example, in the present study, students were asked to write persuasive texts about animals that are difficult to classify. The top-level goal for this task included both a rhetorical aspect (persuasion), and a content aspect (biological taxonomy). In this paper, we will refer to the top-level goal that guides a piece of writing as the "goal," and to the particular goals that are subordinate to this top-level goal, which guide more local writing decisions, as "subgoals."

Content Subgoals, Writing and Learning

Content subgoals refer to the conceptual (i.e., topical) knowledge that the writer could address in a text. Topic knowledge affects planning, text production, and revision (Hayes, 2012; see McCutchen, Teske & Bankston, 2008, for a review). One approach to scaffolding nonfiction writing is to provide students with conceptual subgoals in the form of subtopics to be addressed. For example, The National Council of Teachers of English (n.d.) provides an online resource that

allows students to create an inquiry project on animals; it prompts students to select a graphic organizer that includes animal facts, babies, interactions, and habitats; in turn, each of these organizers embeds prompts to write about specific concepts, e.g., the "Facts" organizer leads to prompts to write sentences about animal appearance, movement, and diet.

It can be theorized that content goal prompts elicit the writer's attention to the conceptual matter of a text. Bereiter and Scardamalia's model of writing (1987; Chuy, Scardamalia & Bereiter, 2012; Kellogg, 2008) has been influential in the literature on writing goals (e.g., Butcher & Kintsch, 2001; McNeill and Krajcik, 2009). This model includes a content problem space in which the writer addresses the question "What do I mean?" and a rhetorical problem space in which the writer addresses the question, "What do I say?" Writers who are relative novices may rely on a *knowledge telling strategy*. In this strategy, the writer retrieves or constructs knowledge in the content space, and then translates this knowledge into a goal in the rhetorical space to express this content. Conversely, more expert writers may rely on a knowledge transforming strategy. This is a dialectical strategy, in which the writer may initially adopt a rhetorical goal. The writer translates this into a content subgoal, which is pursued in the content space, using operations such as inference and decision-making. The resulting proposition may then be translated back into the rhetorical space for expression in writing.

The study that has most directly investigated the effects of content subgoal prompts on writing involved having university students compose introduction sections for psychology articles (Butcher & Kintsch, 2001). First, students studied guidelines for writing an introduction section, as well as background information about the topic of the article, and the text sections that would follow the introduction and comprise the remainder of the article. In a 2 x 2 design, participants received either rhetorical prompts, content prompts, neither type of prompt, or both

types. In the first experiment, participants received the prompts after reading the sources, i.e., immediately prior to writing. The results were complex: Briefly, the content prompts led the writers to invest more time in planning and drafting, and to write holistically better texts. In a second experiment, the prompts were provided prior to reading the source material. Rhetorical prompts somewhat increased recall of source material; content prompts increased recall of source material further, and increased planning and drafting time, and the quality of students' texts. The authors interpreted these results in terms of Bereiter and Scardamalia's (1987) knowledge telling model, proposing that for these writers, who were novices in the science report genre, the content prompts elicited problem solving in the content space and that they translated the content generated into text; the writers were less able to translate the rhetorical goals into content goals and content problem solving (cf., McNeill & Krajcik, 2009).

Whereas Butcher and Kintsch (2001) focused on the introduction to the psychology article. In the present study, the focus was argumentation, because of its role in learning in science (Kuhn, 2010; Osborne, 2010). We hypothesized that content subgoals, compared to a control condition (goal only, with no subgoal prompts), would result in more concepts being addressed, a larger number of complex propositions in text, higher text quality, and higher learning as operationalized by a post-test of content understanding.

Scaffolding Rhetorical Subgoals

"Rhetoric" is defined here as the communicative dimension of a text, that is, its intended effect on the reader. In argumentation, a common type of rhetorical goal is that of persuading the reader (Walton, 1998). The types of rhetorical subgoals that can serve as means to persuasion include the following: giving reasons to support a claim, acknowledging an alternative claim, counter-arguing against the alternative claim or reasons, rebutting a counter-argument to the

writer's claim, and creating a conclusion (e.g., Ferretti et al., 2009; cf., Kuhn, 2010). Many elementary and secondary students create arguments that are brief, comprising only an opinion and a few reasons; they typically omit dialectical aspects of argumentation, such as counter-arguments and rebuttals, and neglect evidence for claims that differ from their own (Felton & Kuhn, 2001; Ferretti et al., 2009; Means & Voss, 1996; Wolfe, Britt & Butler, 2009).

To scaffold elementary writers, several researchers have investigated the effect of providing prompts for rhetorical subgoals. In a comprehensive meta-analysis of experimental research on elementary writing instruction, it was found that prompting "product goals," which frequently comprised what have been called here rhetorical argument subgoals, reliably improved holistic text quality (Graham et al., 2012). For example, Ferretti et al., (2009) asked fifth and sixth grade students to write arguments. Students in an elaborated goal condition received grade-appropriate prompts to address the following: reasons, an alternative claim, reasons for the alternative claim, rebuttals to these reasons, and a conclusion. The elaborated goal group, compared to a group that received only the general argument goal, produced texts that were significantly higher in holistic quality (cf., Golanics & Nussbaum, 2008; Nussbaum & Kardash, 2005).

A question concerns the role of rhetorical subgoals in writing to learn. According to the knowledge transforming model discussed above (Bereiter & Scardamalia, 1987), skilled writers engage in a dialectic between rhetorical problem solving and content problem solving. As noted above, several studies have demonstrated that rhetorical subgoals increase text quality, but very few have investigated the effect of rhetorical subgoals on learning. The knowledge transforming model implies that argumentation subgoals could elicit reasoning about content in several ways:

Evaluating claims on the basis of evidence, considering alternative claims, counter-arguing against alternative claims, and rebutting counter-arguments.

Previous research has shown that the variety of argument moves that students include in text predicts their learning during argument writing (Klein & Samuels, 2010). In a quasi-experimental study with middle school students, McNeill and Krajcik (2009) scaffolded scientific "explanations" (arguments) using either domain general prompts (similar to "rhetorical subgoals" here), or context-specific prompts (similar to "content subgoals" here). They found that context-specific prompts supported explanation writing more effectively than domain general prompts, but only when the teacher also provided domain general support. Other studies have included argument or explanation writing with rhetorical and content goals, but the effects of these two kinds of goals have not been investigated separately (Lee & Songer, 2004; Sandoval & Reiser, 2004).

The results of the two studies discussed above (Butcher & Kintsch, 2001; McNeill & Krajcik, 2009) are similar in providing initial support for the effects of content goals on writing and learning. However, further research is needed. In these studies, the rhetorical goals and content goals were not entirely comparable. For example, Butcher and Kintsch (2001) implemented content goals by providing five specific content prompts, but implemented rhetorical goals by providing two general rhetorical prompts, so the conditions differed with respect to the type of goals, the number of goals, and their specificity.

In order to control for specificity, we make a distinction between goal-setting, which is relatively global and concerns the overall purpose of a text, and subgoal setting, which is more local. By focusing on the subgoal level, it is possible to compare content prompts and rhetorical prompts, which are equal in number, and equally focused on a relatively local level of discourse.

For example, in an argument on the question, “Is a dolphin a fish or a mammal?” the rhetorical subgoal prompt of “giving a reason” in a rhetorical subgoal condition can be paralleled by the content subgoal prompt of “respiration” in the content subgoal condition. In response to either type of prompt, a student could write, “A dolphin is mammal because it breathes with lungs.”

Consequently, based on the notion that rhetorical subgoal prompts can elicit critical reasoning about content, we hypothesized that rhetorical subgoal prompts, compared to a control condition (persuasive goal prompt with no subgoals), would result in a greater variety of rhetorical moves, more frequent complex propositions in text, higher text quality, and higher post-test score on classification.

Writing Subgoals and Cognitive Load

The theory above implies that content subgoals prompt the retrieval of conceptual information from long term memory or text sources relatively directly (Bereiter & Scardamalia, 1987; cf. Kellogg, 2008). This theory is consistent with the findings of previous comparisons of content and rhetorical goals discussed above (Butcher & Kintsch, 2001; McNeill & Krajcik, 2009). The knowledge-transforming model implies that, in contrast, a rhetorical subgoal prompt such as “provide reasons” would only cue the retrieval of content-knowledge indirectly; that is, the learner must first interpret what kind of content could comprise a “reason,” and search for it.

The effects of subgoal setting in learning activities has been extensively theorized and investigated in cognitive load theory (e.g., Sweller, Ayres & Kalyuga, 2011). Cognitive load theory recognizes three sources of working memory load in learning activities. Intrinsic cognitive load is determined by the number of interacting elements that the learner must relate to one another and by the learner’s prior knowledge. Extraneous cognitive load is imposed by inefficient instructional methods, such as unnecessary means-end problem solving. Germane

cognitive load arises from activities that contribute to learning, such as self-explanation. (Currently, an issue is whether germane cognitive load should be conceptualized as a component of intrinsic cognitive load, or as a separate source of cognitive load that combines additively with it, e.g., Kalyuga, 2011).

Previous research has shown that an important source of extraneous cognitive load is means-end problem solving. For example, Sweller (1988) found that mathematics tasks that require means-end problem solving resulted in significantly higher cognitive load and lower learning than goals that did not require means-end problem solving (cf., Sweller, 2010; Wirth, Künsting & Leutner, 2009). Based on the analysis above, we proposed that content subgoal prompts relatively directly elicit content that would support inferences and learning; conversely, rhetorical subgoal prompts would only indirectly elicit content that would support inferences and learning. That is, rhetorical prompts goal setting is based on an additional level of subgoaling, or means-end problem solving. Therefore, we hypothesized that content subgoals would result in lower extraneous cognitive load than rhetorical subgoals, and produce higher post-test scores on classification knowledge.

The Present Study

In the present study, the effects of rhetorical and content subgoals were investigated in the context of writing to learn about the classification of vertebrates. Classification is a common elementary science topic, which provides a foundation for understanding anatomy, physiology, and evolution. Classification lends itself to argumentation as a means of critical thinking due to the prevalence of student misconceptions about this topic (e.g., Braund, 1991; Burgoon & Duran, 2012; Kattmann, 2001). Many students classify animals based on surface features or behaviors rather than essential or internal characteristics; they classify non-hierarchically; and they have

limited knowledge of the characteristics of each class of organisms (e.g., amphibians). These misconceptions lead students to misclassify specific animals: For example, many elementary students believe that dolphins are fish; turtles are amphibians; and penguins are mammals (Prokop, Kubiatko & Fančovičová, 2007; Yen, Yao & Chiu, 2004; Yen, Yao & Mintzes, 2007).

In the current study, students completed argument writing activities about vertebrates that are difficult to classify. Each student was randomly assigned to write in one of three conditions: Students in the argument goal-only (control) condition received a prompt asking them to classify an animal and persuade the reader of their opinion. Students in the rhetorical subgoal prompt condition received the same argument goal prompt, plus six rhetorical subgoal prompts: to include several reasons for the opinion, an alternative claim, reasons for the alternative claim, and a counter-argument to the alternative claim. Students in the content subgoal condition received the argument goal prompt, plus six content subgoal prompts concerning attributes used for classification: respiration; skin or body covering; warm-blooded or cold-blooded; birth; feeding of young; and movement.

To summarize, the hypotheses introduced above were the following:

- First, the content subgoal condition, relative to the argument goal-only (control) condition, would result in higher scores on the following: number of attributes addressed, holistic text quality, number of complex propositions, and post-test learning about classification.
- Second, the rhetorical subgoal condition, relative to the argument goal-only condition, would result in higher scores on the following: variety of rhetorical moves, number of complex propositions, holistic text quality, and post-test learning about classification.

- Third, the content subgoal condition, relative to the rhetorical subgoal condition, would result in lower cognitive load, which would mediate higher post-test scores on learning about classification.

Method

Participants and Curriculum Context

The science topic for this study, vertebrate classification, was based on the Biodiversity unit of the Grade 6 Ontario, Canada curriculum in science. Multi-grade grouping is common in this province, so Grade 5 and Grade 7 students in multi-grade classes also participated in the study. The activities took place at the beginning of the unit so as to avoid the complicating effects of classroom instruction. A letter of information was distributed to parents and students; every student who consented, and whose parent/guardian also consented, was included in the study. Two students, who had consented but did not complete the writing activities due to prolonged absences from school, were deleted from the study.

The final sample comprised 17 Grade 5 students, 36 Grade 6, and 20 Grade 7 students. It included 38 girls and 35 boys. Age appropriate placement is used in this province, so students ranged from 10 years 1 month for Grade 5 students to 13 years, 5 months for Grade 7 students. The distribution of previous writing achievement as defined by most recent report grade was the following: 23 of the students achieved below grade level expectations of a 70% average; 41 were in the expected grade range of 71% to 79%; and 9 were above the expected grade range.

Procedures and Materials

Pretest quiz on the classification of vertebrates. Participants completed a multiple choice pretest of classification knowledge. The purpose of the pretest was to provide a covariate that would increase the power of the hypotheses tests concerning the effects of treatment

conditions on post-test classification scores. Test items were based on previous research concerning classification (e.g., Burgoon et al, 2012; Prokop et al., 2007; Yen et al., 2007). Topics included the characteristics of each class of vertebrates, the classification of particular species, criteria for classification, and hierarchical classification. The test included 15 items, each of which included 4 options. Items were illustrated with colour photographs of organisms.

First writing activity: Classifying dolphins. Approximately five days after pretesting, students completed the first writing activity. Each received a package of three brief documents titled "Mammals," "Fish," and "Dolphins." These documents did not tell students the correct classification of dolphins, but provided information that could allow students to validly infer the correct classification. The source materials used grade-appropriate language; the text was brief; and the materials were extensively illustrated with color photographs.

Each participant was randomly assigned to one of three writing goal prompt conditions: (1) Persuasive goal only (control) condition; (2) persuasive goal + content subgoal prompt condition; or (3) persuasive goal + rhetorical subgoal prompt condition. Students within each class were randomly assigned to conditions using digits ranging from 1 to 3, generated by an electronic random number generator. Consequently, each of the three conditions were approximately equally represented within each class.

Persuasive goal only (control) condition. One-third of the students were randomly assigned the persuasive goal prompt only: "Is a dolphin a fish or a mammal? Give your opinion and persuade the reader."

Content subgoal prompt condition. One-third of the students were randomly assigned the persuasive goal prompt, plus the following content subgoal prompts:

"…What is your opinion? To persuade the reader, please write about…

- How it breathes
- Its skin or covering
- Warm-blooded or cold-blooded
- How it is born
- If it feeds its young, or how it feeds its young
- How it moves

The first five of these prompts provided the opportunity for students to construct valid reasons, e.g., the fact that dolphins are born "alive" rather than hatched from eggs supports the claim that they are mammals. The sixth prompt, "how it moves," provided the opportunity for students to identify the possible alternative claim that a dolphin is a fish, and to counter-argue against it.

Rhetorical subgoal prompt condition. One third of the students were randomly assigned the persuasive goal prompt, plus the rhetorical subgoal prompts, similar to Ferretti et al. (2009):

"…What is your opinion? To persuade the reader, please remember to…

- Give a reason for your opinion.
- Give a second reason.
- Give a third reason.
- Give a fourth reason.
- Give a fifth reason.
- What opinion could someone have that is different from yours? What reason could they give for their opinion? Explain why their reason is not a good one."

It is notable that these rhetorical prompts could be answered by providing reasons referring to the same attributes listed in the content subgoal prompts (e.g., respiration, skin covering, etc.).

Second writing activity: Classifying sea turtles. The second writing activity took place three days later; it was structured identically to the dolphin activity. Students wrote on the question, "Is a sea turtle a reptile or an amphibian?" Each student received information sheets about "Reptiles," "Amphibians," and "Sea Turtles." Each also received the same type of writing prompt that he or she had received in the first activity, i.e., (1) a persuasive goal prompt only; or (2) a persuasive goal prompt plus rhetorical subgoal prompts; or (3) a persuasive goal prompt plus content subgoal prompts.

Transfer writing activity: Classifying the penguin. The purpose of this activity was to assess students' transfer of rhetorical and conceptual elements from the first two writing activities to a novel topic. Five days after the second writing activity, students completed an activity on the question, "Is a penguin a bird or a mammal? What is your opinion? Persuade the reader." All students received information booklets about "Penguins," "Birds," and "Mammals." As with previous writing tasks, the information in the sources was sufficient to allow students to infer the correct classification of the animal, and to construct claim-evidence relationships, but the sources did not provide these relationships for the student. It is important to note that for the transfer task, all students received only the argument goal prompt, with no rhetorical or content subgoal prompts.

Ratings of cognitive load. Immediately after completing each writing activity, students rated their cognitive load. Previous research indicates that Likert scale ratings of effort and perceived difficulty are valid measures of this construct (Brünken, Seufert, & Paas, 2010; DeLeeuw & Mayer, 2008). The first Likert scale asked, "How easy or difficult was this writing activity? Please circle a number." Numbers ranged from "1, Very, very easy" to "9, Very, very difficult."

The second scale asked, “How much effort did you put into this writing activity? Please circle a number.” The scale ranged from “1, Very, very little effort” to “9, Very, very much effort.”

Post-test quiz on classification. Finally, students completed the post-test quiz on classification. The purpose of the post-test score was to allow a between-subjects test of the effect of treatment conditions. The content was based on previous research on classification, and included the following topics: Classification of difficult animals; preference for taxonomic classification; preference for using defining characteristics to classify animals; and the hierarchical nature of classification (e.g., Burgoon et al, 2012; Prokop et al., 2007; Yen et al., 2007). The test included 20 items with a maximum total score of 31 points. All items were multiple choice questions; this format was selected to avoid the problem common in the writing to learn literature, of using post-tests comprised of writing activities similar in form to the experimental treatment (see Hebert et al., 2013 for a review). Each item included 4 options. Items were illustrated with color photographs of vertebrates. Eighteen points were based on two-tiered multiple choice questions: The first tier of each question pair required students to make a judgment concerning classification; the second tier of each pair required students to select a reason to justify this judgment (Treagust, 1988; Yen et al., 2004). Students could receive a point for the first part of a two-tiered question, even if the second part was incorrect. The post-test score was not intended to be directly compared to the pretest score, for example as a gain score; consequently, it was not necessary for the pretest and post-test to have the same total maximum score. Inter-item reliability was high, $\alpha = .83$.

Qualitative Coding

The condition of texts was masked prior to rating and analysis. All texts were rated and coded by two independent reviewers. Text quality was rated by an experienced elementary

teacher unaware of the purpose of the study and a research officer experienced in writing research. Differences in ratings of continuous variables (e.g., text quality) were resolved by averaging; differences in categorical variables (e.g., type of rhetorical move) were resolved by a third rater who was a graduate research assistant.

Validity of classification. For each writing activity, this variable dichotomously coded whether the student classified the organism correctly.

Holistic text quality. To avoid biasing the assessment in favour of any writing condition (e.g., rhetorical subgoals versus content subgoals), a purely holistic rating of text quality was elicited using the question, "How good is this as a persuasive text?" Each rater sorted the texts into nine levels, ranging from 1 for well below average, to 5 for average, to 9 for well above average. Inter-rater reliability for each of the three texts was good (Spearman $r = .75$, $p < .001$).

Complexity of propositions. For each text, each of the attributes (e.g., respiration) listed in the hypotheses was coded with respect to the proposition constructed about it. Levels of complexity included the following: a *complex proposition* connected all three concepts (attribute, family and class), e.g. *a dolphin* [family] *is a mammal* [class] *because it breathes air using lungs* [attribute]; a *simple proposition* connected two of the three concepts of attribute, family, and class; *missing* meant that no proposition was written about a given attribute. For each level of complexity (e.g., "simple"), the number of propositions in a given text was counted, producing a variable that ranged from zero to six propositions. Inter-rater reliability using Spearman correlation were excellent, across the three writing topics and three categories of complexity, it ranged from $r = .72$ to $r = .86$; all $p < .001$.

Rhetorical moves. Each sentence in the text was classified by type of rhetorical move. The categories for rhetorical moves were the following: claim; reason; counter-argument (against

writer's claim); alternative claim; reason for alternative claim; rebuttal of alternative claim or rebuttal of reason for alternative claim; conclusion; other (i.e., not an argument move). As expected, reasons were relatively frequent and the number of reasons per text was normally distributed; inter-rater reliability, $r = .74$, $p < .001$. All other rhetorical operations (e.g., claim, conclusion) were Poisson distributed, so they were coded dichotomously (present versus absent). The number of different *types* of argument moves was summed for each text, yielding a normally distributed variable, *variety of rhetorical moves*, ranging from 0 to 8.

Results

Data Screening and Assumption Testing

Data was screened for consistency with the assumptions of multivariate analysis of covariance. Three variables were positively skewed, so each was normalized using a square root or square root (x + 1) transformation. Two participants comprised univariate outliers, and one participant comprised a multivariate outlier; MANCOVA is sensitive to outliers, so these cases were deleted from the analysis. Because the number of participants in each class was not evenly divisible by three, the number of participants was 24 in the general goal condition, 25 in the content subgoal condition, and 24 in the rhetorical subgoal condition. With these adjustments, all assumptions of MANCOVA were met. Additionally, five participants were missing one or more data points, most often a Likert scale rating of effort or difficulty. To avoid listwise deletion of these cases during MANCOVA, values were imputed using the SPSS missing data regression procedure. Imputed values comprised less than 1% of the total data.

In the MANCOVA for each writing activity, the independent variable was writing subgoal condition. The covariates comprised previous writing achievement and classification pretest score. The dependent variables were effort, perceived difficulty, text quality, variety of

rhetorical moves in text, number of complex propositions, number of simple propositions, and number of reasons. In an initial analysis, grade level did not affect the dependent variables, so it was omitted from subsequent analyses. In the presentation below, the terms small, medium and large are used to characterize magnitude of effects; these terms should only be considered approximations. Because classification accuracy was a dichotomous variable, it was analyzed using chi-squared tests.

Pretest

The three writing conditions did not differ significantly with respect to the classification pretest, F (2, 70) = .13; persuasive goal condition M = 11.58, SD = 2.19; content subgoal condition M = 11.88, SD = 2.19; and rhetorical subgoal condition M = 11.88, SD = 2.58. The groups also did not differ significantly with respect to grade level, F (2, 70) = .08; nor previous writing achievement F (2, 70) = 1.04.

Results of First Writing Activity: Classifying the Dolphin

Classification accuracy showed a statistically significant difference across conditions, X^2 (2, 73) = 9.00, p = .01; 23 students in the rhetorical condition, 18 students in the goal-only condition, and 19 students in the content subgoal condition inferred the correct classification in their written text.

Writing condition had a statistically significant, large effect on the combined dependent variable, Wilk's Λ = .45, F (14, 124) = 4.34, p < .001, partial η^2 = .33. The classification pretest covariate did not account for significant variance in the combined dependent variable, Wilk's Λ = .81, F (7, 62) = 2.08, p = .06, partial η^2 = .19. The previous writing achievement covariate did not significantly predict the combined dependent variable, Wilk's Λ = .85, F (7, 62) = 1.59, p

=.16, partial η^2 = .15. See Table 1 for the descriptive statistics and the results of Bonferroni post hoc comparisons among the writing conditions for the Dolphin activity.

Writing condition had a statistically significant effect, which was small in magnitude, on the quality of written texts, F (2, 68) = 3.27, p = .04, partial η^2 = 09; quality was significantly higher in the rhetorical subgoal condition than the goal-only condition. The perceived difficulty of the writing task did not differ significantly by writing condition, F (2, 68) = 1.25, p = .30, partial η^2 = 04. Reported effort also did not differ significantly across conditions, F (2, 68) = 1.40, p = .25, partial η^2 = .04.

Writing condition had a statistically significant, large effect on the variety of rhetorical moves in text, F (2, 68) = 9.80, p < .001, partial η^2 = .22; it was significantly higher in the rhetorical subgoal and goal-only conditions than the content subgoal condition. The number of reasons did not differ significantly across conditions, F (2, 68) = 1.40, p = .25, partial η^2 = .04. Writing condition had a statistically significant, large effect on the number of simple propositions in text, F (2, 68) = 17.82, p <.001, partial η^2 = .34, such that it was higher in the content subgoal condition than the other two conditions. The number of complex propositions showed a statistically significant, moderate difference across conditions, F (2, 68) = 4.65, p = .01, partial η^2 = .12, such that it was higher in the rhetorical subgoal condition than the content subgoal condition.

Results of the Second Writing Activity: Classifying the Sea Turtle.

Writing condition had a statistically significant effect on the number of students who classified the sea turtle correctly during writing, X^2 (2, 73) = 4.96, p = .04; 21 students in the rhetorical subgoal condition, 17 students in the goal-only condition, and 15 students in the content subgoal condition made correct classifications.

Writing condition had a statistically significant, large effect on the combined dependent variable, Wilk's Λ = .32, F (14, 124) = 4.91, p < .001, partial η^2 = .31. The pretest classification knowledge covariate did not account for significant variance in the combined dependent variable, Wilk's Λ = .95, F (7, 62) = .44, p = .87, partial η^2 = .05. The previous writing achievement covariate did not significantly predict the combined dependent variable, Wilk's Λ = .90, F (9, 57) = .98, p =.45, partial η^2 = .10. See Table 2 for descriptive statistics and Bonferroni comparisons for the Turtle activity.

Writing condition had a statistically significant, medium-sized effect on the quality of written texts, F (2, 68) = 4.16, p = .02, partial η^2 = .11; texts in the rhetorical subgoal condition were rated significantly higher than those in the content subgoal condition. Writing condition did not significantly affect the perceived difficulty of the writing task, F (2, 68) = 1.83, p = .17, partial η^2 = 05; nor did it affect reported effort, F (2, 68) = 1.15, p = .32, partial η^2 = .03.

Writing condition had a statistically significant, large effect on the variety of rhetorical moves, F (2, 68) = 13.97, p < .001, partial η^2 = .29, such that it was significantly higher in the rhetorical subgoal condition and goal-only condition than in the content subgoal condition. The number of reasons did not differ significantly by condition, F (2, 68) = .52, p = .60, partial η^2 = .02. Writing condition had a statistically significant, large effect on the number of simple propositions in text, F (2, 68) = 25.91, p <. 001, partial η^2 = .43; it was higher in the content subgoal condition than the two other conditions. The number of complex propositions showed a statistically significant, small difference across conditions, F (2, 68) = 3.51, p =.04, partial η^2 = .09, such that it was significantly higher in the rhetorical subgoal and goal-only conditions, than the content subgoal condition.

Results of Transfer Writing Activity: Classifying the Penguin.

Recall that in the transfer activity, all students received the general persuasive goal prompt only. The number of valid classification inferences did not differ significantly across conditions, X^2 (2, 73) = 1.32, p = .26; the correct classification was made by 22 students in the rhetorical subgoal condition, 21 students in the goal-only condition, and 21 students in the content subgoal condition. The MANCOVA showed that writing condition did not significantly affect the combined dependent variable, Wilk's Λ = .730, F (14, 124) = 1.51, p = .11, partial η^2 = .15. See Table 3 for descriptive statistics for the Penguin activity. The pretest classification covariate did not significantly affect the combined dependent variable, Wilk's Λ = .974, F (7, 62) = .23, p = .98, partial η^2 = .03. The previous writing achievement covariate also did not significantly affect the combined dependent variable, Wilk's Λ = .915, F (7, 62) = .82, p =.57, partial η^2 = .09.

Post-test of Classification

An overall F test did not show a statistically significant effect of condition on post-test classification score, F (2, 68) = 2.91, p = .06, partial η^2 = .08. The planned comparison between the rhetorical condition and the control condition was statistically significant and small to medium in size, CE = 2.92, SE = 1.26, p = .02. The planned comparison between the content subgoal condition and the goal-only condition was not statistically significant, contrast estimate = 2.15, SE= 1.24, p = .08. The pretest classification score strongly predicted the post-test classification score, F (1, 68) = 24.95, p < .001, partial η^2 = .27. Prior writing achievement did not significantly affect classification post-test score, F (1, 68) = 1.63, p =. 20, partial η^2 = .02.

Path Analysis

The treatment conditions did not significantly affect the measures of cognitive load, ruling out cognitive load as a variable that could mediate the relationship between writing

conditions and classification post-test knowledge. Consequently, a path analysis was conducted to investigate the possible mediating role of text variables indicative of argumentative reasoning. The treatment conditions were considered exogenous variables; the text measures were considered possibly mediating variables; and post-test classification score was considered the final criterion variable.

Prior to the analysis, the writing subgoal conditions were coded as indicator variables: A content subgoal indicator variable was created in which the content subgoal condition was coded as 1 and the goal-only (control) condition was coded as 0; and a rhetorical subgoal indicator variable was created in which the rhetorical subgoal condition was coded as 1, and the goal-only (control) condition was coded as 0.

To keep the total number of parameters relative to the number of participants manageable, pairs of corresponding variables across the two writing activities that correlated (e.g., holistic quality of the turtle text and holistic quality of the dolphin text) were collapsed by being summed together. Then, a principle components analysis was used to reduce three of mediating variables (variety of rhetorical moves; number of complex propositions; and holistic text quality) to a single variable. Only the first component had an Eigen value greater than 1; it accounted for 74% of the total variance. On this component, holistic argument quality had a loading of .90; variety of rhetorical moves had a loading of .86; and number of complex propositions had a loading of .82. This component was dubbed the "rhetorical quality factor." The other possible mediating variable to be tested for entry into the path analysis was the number of simple propositions. This reduced set of variables allowed a ratio of approximately 12 participants per parameter. This was considered acceptable, given that the parameter estimates were medium in size (Wolf, Harrington, Clark, Miller, 2013).

For the path analysis, relationships were selected almost entirely empirically. For the first regression, the criterion variable was post-test classification; all other variables in the model were evaluated using a backward elimination algorithm, with a criterion of $p < .05$ for entry and $p \geq .10$ for removal. For each subsequent regression, an endogenous variable was set as the criterion variable, and the remaining variables were evaluated as predictors.

See Figure 1 for the resulting path diagram. All relationships were significant at $p < .01$. The magnitude of most relationships was medium in size; partial correlations ranged from small ($\beta = .27$) to large ($\beta = .68$). See Table 4 for a comparison of the observed and implied correlations. Of the 10 implied correlations, 8 were within .05 of the observed correlation; the other 2 implied correlations were within 0.10 of the observed correlations.

For a summary of the causal effects, see Table 5. The model accounted for 23% of the variance in post-test classification score, which was the final endogenous variable. Although the amount of variance accounted for in post-test classification was modest, recall that the purpose of the path analysis was specifically to test whether argument text characteristics mediated the effect of the writing conditions on the post-test classification score. The determinants of post-test classification comprised rhetorical text quality, which was medium in magnitude, and simple propositions, rhetorical subgoal condition, and content subgoal condition, which were small in magnitude. The model accounted for 27% of the variance in rhetorical text quality; rhetorical subgoal condition was a positive determinant of rhetorical text quality; the number of simple propositions and the content subgoal condition, were both negative determinants. The model accounted for 46% of the variance in the number of simple propositions; this was strongly determined by the content subgoal condition.

Discussion

This was the first randomized experimental study to investigate the effects of content subgoal prompts and rhetorical subgoal prompts on text quality and learning in elementary science argument writing. In these writing to learn activities, the source documents did not inform students of the correct classification of each vertebrate. Rather, the sources provided information about classes, families and attributes of vertebrates; students were required to infer the correct classification, and to compose an original argument by constructing valid relationships among these concepts. The results included many statistically significant effects, most of which were replicated across the two writing to learn activities; however, several effects differed from the initial hypotheses.

How Do Content Subgoals Affect Writing?

The first hypothesis was that content subgoal prompting, compared to goal-only prompting, would result in higher scores on the dependent variables. Descriptively, the average text in the content subgoal condition included a claim and three reasons, with no dialectical argument moves. The main positive effect of content prompts was an increase in the number of simple propositions that related two science concepts to one another. This means that the content condition resulted in students addressing a broader set of the key concepts than the other two conditions. However, the content subgoal condition also resulted in a significantly lower variety of rhetorical moves and fewer complex propositions that connected three or more science concepts, than the control condition and the rhetorical subgoal condition. The writers in the content subgoal condition offered a similar number of total reasons to writers in the other conditions; however, they typically did this by stating one reason in a fully explicit way, e.g., “A turtle is reptile because it hatches from eggs on land;” they then expressed additional reasons using simple propositions, leaving their rhetorical status as reasons implicit, e.g., “Also, young

turtles breathe with lungs," omitting that this was a characteristic of reptiles. Unlike rhetorical goal prompting, content goal prompting did not have significant effect on learning.

Why did content prompting produce this pattern of effects? The results are consistent with the theory that during writing, content goals lead to the retrieval of relevant concepts from long term memory or selection of relevant concepts from source documents (Bereiter & Scardamalia, 1987; Butcher & Kintsch, 2001). Retrieval made this information available for inferences and claim-evidence relationships. However, the content prompts appeared to have partially distracted students from the rhetorical aspects of composing the text. This was evident in the fact that students in this condition employed less variety of rhetorical moves than those in the control group.

These results are somewhat different from previous studies of the effects of content prompts, which found that they significantly increased text quality relative to rhetorical goals (Butcher and Kintsch, 2001; McNeill & Krajcik, 2009). Each of these former studies differed from the present one in one of several ways: educational level of the participants, text genre, science content, or time frame. Perhaps the most notable difference is that both of the former studies contrasted relatively specific content goals versus more abstract rhetorical goals. In the present study, the rhetorical subgoals and content subgoals were parallel, in both number and specificity. It is not claimed here that these previous studies presented methodological problems, or that this difference in findings demonstrates that rhetorical subgoals are more effective than content subgoals. Rather, we suggest that it is necessary to consider both the type of goal prompt, and its level of specificity, in predicting its effects on writing and learning.

How did Rhetorical Goals Affect Writing?

The second hypothesis was that the rhetorical subgoal condition, relative to the control condition, would result in higher scores on the dependent measures; this was partially supported. Descriptively, the average student in the rhetorical subgoal condition produced a text with three different kinds of rhetorical moves, including a claim, four reasons for the claim, and one dialectical move such as a counter-argument. The rhetorical subgoal condition, relative to the goal-only control condition, resulted in a greater frequency of valid classification inferences during the first two writing activities, higher text quality on the first writing activity, and higher scores on the classification post-test. Relative to the content subgoal condition, the rhetorical condition resulted in a greater variety of rhetorical moves and a greater number of complex propositions, and on the second writing activity, it resulted in greater holistic text quality.

To our knowledge, the present study is the first to show that rhetorical subgoal prompts increase learning, relative to a goal prompt alone. The present results also confirm previous research in showing that rhetorical subgoal prompting improves text quality, and extends this finding to science content (Ferretti et al., 2009; Golanics & Nussbaum, 2008; Goldstein, Crowell & Kuhn, 2009). Educationally, these results suggest that in classes where students have not yet mastered dialectical argumentation, argument tasks can be made more effective for learning by providing students with rhetorical subgoal prompts.

What Mediates Learning During Argument Writing?

The third hypothesis was that rhetorical and content goals would differ in their effects on cognitive load, which would in turn affect learning. This was based on the assumption that rhetorical subgoals elicit extraneous cognitive load by requiring writers to engage in an additional level of means-end analysis by translating rhetorical subgoals into content subgoals. However, the type of writing condition did not significantly affect either measure of cognitive

load. Rather, participants rated all three conditions as moderate in difficulty; that is, the average ratings were within one standard deviation of the midpoint of the scale, which was labelled "medium." Participants identified their own level of effort, on average, approximately one standard deviation above the midpoint of the scale.

Given that cognitive load theory has only recently been applied to writing, these results comprise useful information. Previous research has shown that tasks which demand subgoaling (i.e., means-end analysis) produce higher cognitive load than those which do not require means-end analysis (Sweller, 1988). In this context, the present results suggest that the three conditions may not have differed in demand for means-end problem solving during writing. A plausible interpretation, consistent with the results of Butcher & Kintsch (2001, Experiment 2) is that because students are provided with rhetorical subgoals *prior* to writing, they may have selected relevant content during initial reading of the sources, reducing the need to later derive content subgoals from rhetorical subgoals during the writing phase (cf., Klein, 2014).

The path analysis suggests that rhetorical subgoal prompts and content subgoal prompts elicit two somewhat different paths to text quality and learning (Figure 1). One path is that which has been identified in a longstanding conception of writing to learn: The call for argumentation presents a rhetorical goal; writers pursue this goal by constructing complex propositions, and this contributes to learning (Applebee, 1984; Bereiter & Scardamalia, 1987; See Klein, 1999 for a review). This theory was supported by the path in which rhetorical subgoal prompts increased the rhetorical quality of text. Recall that rhetorical text quality was a factor that was loaded with three variables: variety of rhetorical moves, holistic text quality, and number of complex propositions. This text factor can be considered a proxy for argumentative reasoning while writing. This factor in turn predicted post-test classification understanding.

The second learning mechanism suggested by the path analysis could be called the content path: Content subgoals elicit attention to relevant concepts, making them available for inferencing and construction of propositions, which contributes to learning (Figure 1). This was apparent in the path in which content subgoal prompts increased the number of simple (two-concept) propositions, which in turn accounted for unique variance in learning. This content subgoal path to learning is largely distinct from the rhetorical subgoal path discussed in the previous paragraph, not only because of the difference in the conditions that elicited it, but also because of the negative correlation that cross-linked these paths: the number of simple proposition predicted negative variance in the rhetorical quality factor.

Limitations and Further Research

A result that requires further comment is performance on the transfer (penguin) activity. Although the subgoal conditions differed in their effects on the first two writing activities, they did not differ in their effects on the transfer writing activity. These results are similar to those obtained by Schworm and Renkl (2007). They found that in learning about argumentation, students who were prompted to self-explain the learning domain (argumentation principles) learned more about argumentation and transferred their argumentation skills; students who were prompted to self-explain the exemplifying domain (conceptual content) of the argument writing task did not. The students in the present study carried out a writing task that was more similar to self-explaining the exemplifying domain (classification), rather than the learning domain (argumentation) itself. This suggests that students might have performed better on argumentation transfer tasks if they were presented with reflective questions about argumentation principles.

Perhaps the most important limitation of this study was the focus on one genre: argumentation. Argumentation, by its nature, foregrounds rhetorical goals. It would be valuable

to replicate this research in a genre such as a causal explanation or an informational report, in which the top-level goal of the text foregrounds conceptual rather than rhetorical demands. A second limitation of this study is that it focused on one topic, taxonomic classification; further research should investigate whether similar results occur with other science topics, as well as topics in other domains, such as history.

A third limitation is that the measures of cognitive load were based on two Likert scales: effort and perceived difficulty. Although commonly used and well-validated as measures of cognitive load in general, it remains unclear whether these scales measure different components of total cognitive load, and if so, which scales measure which components (DeLeeuw & Mayer, 2008; Leppink, Paas, Van der Vleuten, Van Gog, & Van Merriënboer, 2013). Finally, the number of participants (N = 73) was modest for a path analysis. The size of the parameters was medium, increasing the reliability of the analysis; however, it would be preferable to have a larger sample, with multiple measures per variable (Wolf, Harrington, Clark, Miller, 2013).

Conclusions

This is one of the first studies of the effects of rhetorical and content argument subgoal prompts on writing to learn, so it should be considered exploratory. The results suggest that argument writing in science can be supported by rhetorical or content subgoals, and that these subgoals elicit different paths to text quality and learning. Pending further investigation, these results suggest the following educational implications:

- Teachers could consider assigning brief argument writing activities to help students learn about classification.
- Teachers could consider supporting content area argument writing with rhetorical subgoals in order to increase text quality and learning.

- Teachers could consider assigning content subgoals to elicit engagement with a range of concepts.

References

Applebee, A. (1984). Writing and reasoning. *Review of Educational Research, 54,* 577–596.

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

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.

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

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Braund, M. (1991). Children's ideas in classifying animals. *Journal of Biological Education, 25,* 103-110.

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 University Press.

Burgoon, J. N., & Duran, E. (2012). Investigating elementary teachers' conceptions of animal classification. *School Science and Mathematics, 112,* 410-419.

Butcher, K. R. & Kintsch, W. (2001). Support of content and rhetorical processes of writing: Effects on the writing process and the written product. *Cognition and Instruction, 19,* 277-322.

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.

Chen, C. H., & She, H. C. (2012). The impact of recurrent on-line synchronous scientific argumentation on students' argumentation and conceptual change. *Educational Technology & Society*, *15*, 197-210.

Chuy, M., Scardamalia, M., & Bereiter, C. (2012). Development of ideational writing through knowledge building. In E. L. L. Grigorenko, E. Mambrino, & D. D. D. Preiss (Eds.) *Writing: A Mosaic of New Perspectives* (pp. 175-190). Psychology Press.

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.

Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes*, *32*(2-3), 135-153.

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

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

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

Goldstein, M., Crowell, A., & Kuhn, D. (2009). What constitutes skilled argumentation and how does it develop? *Informal Logic, 29,* 379-395.

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.

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.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29,* 369-388.

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.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23,* 1-19.

Kattmann, U. (2001). Aquatics, flyers, creepers and terrestrials—students' conceptions of animal classification. *Journal of Biological Education, 35,* 141-147.

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of writing research, 1*(1), 1-26.

Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review, 11,* 203-270.

Klein, P. D. (2014). Knowledge construction in collaborative science writing: Strategic simplicity, distributed complexity, and explanatory sophistication. In G. Rijlaarsdam (Series Ed.), & P. D. Klein, P. Boscolo, L. C. Kirkpatrick, & C. Gelati (Eds.), *Studies in*

Writing: Vol. 28, Writing as a Learning Activity (pp. 300-326). Leiden, The Netherlands: Brill.

Klein, P. D., & Kirkpatrick, L. C. (2010). A framework for content area writing: Mediators and moderators. *Journal of Writing Research, 2*(1), 1-54.

Klein, P. D., & Rose, M. A. (2010).Teaching argument and explanation to prepare junior students for writing to learn. *Reading Research Quarterly, 45*, 433-461.

Klein, P. D., & Samuels, B. M. (2010). Learning about plate tectonics through argument writing. *Alberta Journal of Educational Research, 56,* 196-217.

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

Lee, H. S., & Songer, N. B. (2004). Expanding an understanding of scaffolding theory using an inquiry-fostering science program. Retrieved from http://www.biokids.umich.edu/papers/56LeeSongerScaffolding.pdf April 10, 2015.

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior research methods*, *45*, 1058-1072.

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. 451-470). 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.

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

National Council of Teachers of English (n.d.). "Animal Inquiry." http://www.readwritethink.org/files/resources/interactives/animal-inquiry/. Retrieved April 10, 2015.

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., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology, 28*, 384-395.

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.

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

Prokop, P., Kubiatko, M., & Fančovičová, J. (2007). Why do cocks crow? Children's concepts about birds. *Research in Science Education, 37,* 393-405.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education, 88*, 345-372.

Schwarz, B. B. (2009). Argumentation and learning. In N. M. Mirza & A.N. Perret-Clermont (Eds.), *Argumentation and education* (pp. 91-126). Boston, MA: Springer-Verlag US.

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99,* 285-296.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22,* 123-138.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education, 10,* 159-169.

Walton, D. N. (1998). *The new dialectic: Conversational contexts of argument*. Toronto, Canada: University of Toronto Press.

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.

Wirth, J., Künsting, J., & Leutner, D. (2009). The impact of goal specificity and goal type on learning outcome and cognitive load. *Computers in Human Behavior, 25,* 299-305.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73,* 913-934.

Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication, 26,* 183-209.

Yen, C. F., Yao, T. W., & Chiu, Y. C. (2004). Alternative conceptions in animal classification focusing on amphibians and reptiles: A cross-age study. *International Journal of Science and Mathematics Education, 2*, 159-174.

Yen, C. F., Yao, T. W. & Mintzes, J. J. (2007). Taiwanese students' alternative conceptions of animal biodiversity. *International Journal of Science Education*, *29*, 535-553.

Table 1,

Dolphin Writing Activity, Descriptive Statistics and Bonferroni Comparisons

	Condition					
	Goal Only		Content		Rhetorical	
	Mn (SD)		Mn (SD)		Mn (SD)	
Holistic Text Quality	4.47	(1.09)	4.74	(1.51)	5.30	(1.10)
Perceived Difficulty	4.04	(1.46)	4.08	(1.47)	3.33	(1.99)
Effort	6.75	(1.39)	7.36	(1.41)	6.50	(2.00)
Reasons	3.41	(1.68)	3.46	(2.28)	4.22	(1.02)
Variety of Rhetorical Moves	2.59	(1.28)	1.72	(1.24)	3.20	(1.18)
Simple Propositions	1.68	(1.50)	4.16	(1.77)	2.22	(1.34)
Complex Propositions	1.20	(.96)	.82	(1.00)	1.62	(1.20)
	Adjusted Mean (Standard Error)					
Holistic Text Quality	4.47_{a}	(.25)	4.69_{ab}	(.25)	5.36_{b}	(.26)
Perceived Difficulty	4.00	(.33)	4.06	(.33)	3.39	(.33)
Effort	6.74	(.33)	7.31	(.33)	6.56	(.33)
Variety of rhetorical Moves	2.60_{b}	(.25)	1.68_{a}	(.25)	3.24_{b}	(.25)
Reasons	3.45	(.35)	3.46	(.35)	4.18	(.35)
Simple Propositions	1.70_{a}	(.32)	4.19_{b}	(.31)	2.17_{a}	(.32)
† Complex Propositions	$.97_{ab}$	(.12)	$.65_{a}$	(.12)	1.16_{b}	(.12)

† A square root transformation was applied prior to analysis.

* Bonferroni corrected multiple comparison; means that do not share a common subscript differ significantly, $p < .05$

Table 2,

Turtle Writing Activity, Descriptive Statistics and Bonferroni Comparisons

	Condition					
	Goal Only		Content		Rhetorical	
	Mn (SD)		Mn (SD)		Mn (SD)	
Holistic Text Quality	4.77	(1.32)	4.44	(1.23)	5.38	(1.31)
Perceived Difficulty	4.83	(1.93)	3.88	(1.51)	4.08	(2.00)
Effort	6.58	(1.25)	7.12	(1.48)	6.42	(1.74)
Reasons	3.46	(1.16)	3.16	(2.52)	3.67	(1.32)
Variety of Rhetorical Moves	2.82	(.92)	1.60	(1.29)	3.25	(1.22)
Simple Propositions	1.14	(1.19)	3.94	(2.06)	1.08	(1.33)
Complex Propositions	1.37	(1.13)	.72	(1.13)	1.54	(1.50)
	Adjusted Mean (Standard Error) *					
Holistic Text Quality	4.78_{ab}	(.26)	4.38_{a}	(.26)	5.43_{b}	(.26)
Perceived Difficulty	4.83	(.38)	3.86	(.37)	4.11	(.38)
Effort	6.59	(.31)	7.08	(.30)	6.45	(.31)
Variety of rhetorical Moves	2.83_{b}	(.24)	1.57_{a}	(.23)	3.28_{b}	(.24)
Reasons	3.47	(.37)	3.15	(.36)	3.68	(.37)
Simple Propositions	1.11_{a}	(.32)	3.94_{b}	(.32)	1.11_{a}	(.33)
† Complex Propositions	1.50_{b}	(.08)	1.25_{a}	(.08)	1.54_{b}	(.08)

† A square root (x + 1) transformation was applied prior to analysis.

* Bonferroni corrected multiple comparison; means that do not share a common subscript differ significantly, $p < .05$.

Table 3,

Transfer Writing Activity, Descriptive Statistics and Bonferroni Comparisons

	Subgoal Condition					
	Goal Only		Content		Rhetorical	
	Mn (SD)		Mn (SD)		Mn (SD)	
Holistic Text Quality	4.83	(1.09)	4.59	(1.29)	4.80	(1.33)
Perceived Difficulty	3.88	(1.57)	3.48	(1.92)	3.46	(2.06)
Effort	6.50	(1.59)	6.64	(1.71)	5.42	(2.00)
Reasons	3.44	(1.47)	3.30	(1.87)	3.16	(1.34)
Variety of Rhetorical Moves	2.79	(1.14)	2.46	(1.00)	2.91	(.72)
Simple Propositions	.65	(.98)	1.59	(1.36)	1.14	(1.04)
Complex Propositions	1.81	(1.25)	1.34	(1.02)	1.63	(1.27)
	Adjusted Mean (Standard Error)					
Holistic Text Quality	4.84	(.25)	4.55	(.25)	4.85	(.25)
Perceived Difficulty	3.87	(.38)	3.41	(.37)	3.54	(.38)
Effort	6.50_{ab}	(.37)	6.67_{b}	(.36)	5.39_{a}	(.37)
Variety of rhetorical Moves	2.80	(.20)	2.43	(.19)	2.94	(.20)
Reasons	3.43	(.33)	3.26	(.32)	3.21	(.33)
† Simple Propositions	1.24_{a}	(.08)	1.56_{b}	(.08)	1.42_{ab}	(.08)
Complex Propositions	1.82	(.24)	1.30	(.24)	1.67	(.24)

† A square root (x +1) transformation was applied prior to analysis.

* Bonferroni corrected multiple comparison; means that do not share a common subscript differ significantly, $p < .05$

Table 4

Observed and Reproduced Correlations for Path Model

	z1	z2	z3	z4	z5
z1. Content Subgoal Condition		-.51	.68	-.38	.10
z2. Rhetorical Subgoal Condition			-.30	.38	.16
z3. Simple Propositions				-.45	.14
z4. Rhetorical Text Quality					.35
z5. Classification Post-Test					
z1. Content Subgoal Condition		-.51	.68	-.39	.05
z2. Rhetorical Subgoal Condition			-.35	.40	.09
z3. Simple Propositions				-.46	.22
z4. Rhetorical Text Quality					.35
z5. Classification Post-Test					

Table 5,

Summary of Causal Effects for Path Model

		Model Causal Effects		
Outcome	Determinant	Direct	Indirect	Total
Post-test classification	Rhetorical Text Quality	.51	--	.51
(R^2 = .23)	Simple Propositions	.36	-.19	.17
	Rhetorical Subgoal Condition	--	.14	.14
	Content Subgoal Condition	--	.12	.12
Rhetorical Text quality	Simple Propositions	-.37	--	-.37
(R^2 = .27)	Rhetorical Subgoal Condition	.27	--	.27
	Content Subgoal Condition	--	.25	-.25
Simple Propositions	Rhetorical Subgoal Condition	--	--	.00
(R^2 = .46)	Content Subgoal Condition	.68	--	.68