A Framework for Content Area Writing: Mediators and Moderators

Perry D. Klein & Lori C. Kirkpatrick

The University of Western Ontario | Canada

Abstract: Writing can be a tool for communicating and learning in content area subjects. This pretest-posttest quasi-experiment examined the effects of instruction in a content area writing framework on students' text quality and ability to use writing to learn. It also examined the effects of possible moderator variables (gender, previous writing achievement) and mediator variables (genre knowledge, approach to writing). A multilevel analysis was conducted with students nested within classes. Instruction significantly increased argument genre knowledge and explanation text quality, but not argument text quality, explanation genre knowledge, or learning during writing. Gender predicted previous writing achievement and posttest argument text quality, but did not interact significantly with instruction. Previous writing achievement strongly affected several posttest measures, but did not interact significantly with instruction. A path analysis supported the theory that instruction affects genre knowledge, which affects text quality, which predicts learning during writing.

Keywords: Writing skills; writing to learn; individual differences; teaching methods; genre

journal of
WRITING RESEARCH

Klein, P. D., & Kirkpatrick, L. C. (2010). A framework for content area writing: Mediators and moderators, *Journal of Writing Research, 2* (1), 1-46. http://dx.doi.org/10.17239/jowr-2010.02.01.1
Contact and copyright: Earli | Perry D. Klein, Faculty of Education, The University of Western Ontario, 1137 Western Road, London, Canada, N6G 1G7 | pklein@uwo.ca.

Each year, students are asked to write in content area subjects such as science and social studies. Such writing may be intended to serve a variety of purposes: Teaching students to communicate about subject matter; teaching them to write nonfiction texts; providing a balanced curriculum; motivating students whose interests lie in the content areas rather than in literacy; and even increasing content area learning (Duke & Bennett-Armistead 2003; Moss, 2005; Wallace, Hand, & Prain, 2004). However, nonfiction writing is a challenge for many students. For example, in the recent National Assessment of Educational Progress in the United States, only 67% of Grade 8 students scored "sufficient" or higher on writing informative text, and only 60% of Grade 12 students scored "sufficient" on writing persuasive text, although they were required to write only about familiar topics (Salahu-Din, Persky, & Miller, 2008). Moreover, some research suggests that in order to learn from content area writing, students need to be skilled writers or high achievers (Bereiter & Scadamalia, 1987; Rivard, 2004). To date, several studies have focused on *teaching* students how to write in the content areas (e.g., Purcell-Gates, Duke, & Martineau, 2007); other studies have *assigned* writing to support content area learning (Bangert-Drowns et al, 2004; Boscolo & Mason, 2001; Wallace, Hand, & Prain, 2004; for reviews, see Klein, 1999; Tynjälä, 2001). However, we are not aware of any studies that have taught students writing *in order to* increase their ability to use writing as a tool for content area learning.

The purpose of this project was to improve students' writing and learning in content area subjects. This paper attempts to address three inter-related questions: First, does this framework significantly affect content area writing, including text quality and content learning during writing? Second, is this framework effective for a variety of students, including students of both genders, and various levels of achievement? And third, if the framework is effective, what mechanisms, in terms of student knowledge and attitudes, underpin its operation?

1. Toward A Theory of Analytical Writing as Content Area Communication and Learning

This framework is based on the family of cognitive theories that model writing as a problem solving process (Hayes, 1996; Spivey, 1997; see Alamargot & Chanquoy, 2001 for a review), primarily the knowledge transforming model (Bereiter & Scardamalia, 1987). Such models propose that skilled writers and readers possess schemata that represent various text genres (Hayes, 1996; Kintsch & Van Dijk, 1978). These schemata represent the structure of such texts, that is, their components, the relationships among them, and some typical linguistic markers of such relationships (Hayes, 1996; Meyer & Poon, 2001). They may serve a heuristic function during writing (Coe, 1994; Hayes, 1996; Spivey, 1997). For example, an argument scheme may include components such as claim, evidence, and rebuttal; the creation of such components can comprise subgoals for writing.

To fulfill these subgoals, writers would need to set content goals (Bereiter & Scardamalia, 1987). If the writer has content knowledge that is organized according to

the structural relationships appropriate to a given text, then it can be retrieved automatically from long term memory (Bereiter & Scardamalia, 1987; Torrance, Thomas & Robinson, 1996); this would not require any new learning. However, writers frequently do not have knowledge that is organized according to the genre; if this is the case, then they must engage in a more effortful process of transforming knowledge (Bereiter & Scardamalia, 1987). The willingness to construct new knowledge may depend on writer attitudes: Writers could be most willing to engage in this process if they have a deep approach to writing, that is, if they conceive of writing as a process of making meaning (Lavelle, 2007). To construct knowledge, writers search internal sources (long term memory), or external sources (e.g., texts), and select relevant information (Bereiter & Scardamalia, 1987; Hillocks, 2005; Klein, Boman, & Prince, 2007; Spivey, 1997). This information may then be transformed to construct the necessary genre relationships, using operations such as drawing inferences and making decisions (Bereiter & Scardamalia, 1987). The construction of new relationships among ideas for the reader also comprises new learning for the writer. This model will be illustrated below with reference to specific text genres.

Based on this model, a framework for content area writing education was developed in a previous design experiment (Klein & Rose, in review); it is summarized here in Table 1. It is a "framework" in the sense that it is a set of design principles that teachers can use to create writing intensive content area units for a variety of topics in science and social studies; it is not a programme of prepared lessons. It attempts to coherently address several levels and aspects of content area writing education, including programme organization, teaching strategies, activity type, motivational methods, and assessment. It will be described briefly here.

The largest element in the framework is the content area unit of study, in which students write several times each week (Bangert-Drowns et al., 2004). Instruction focuses primarily on two text genres: arguments, which afford critical thinking; and explanations, which afford theoretical understanding. Within a given unit of study, there occur several inquiry writing sequences of three to five sessions (Hillocks, 2005): First, students generate information through an experience, such as reviewing paintings from a historical period; then the teacher presents a writing strategy lesson that can be used to interpret this information, such as incorporating evidence into an argument; then students use this strategy to write a text that interprets the information. Nested within the inquiry lesson sequence, the predominant approach to writing education is cognitive strategy instruction (Graham, 2006); teachers initially take principal responsibility for discussing a strategy with students and leading them in writing, and then gradually release responsibility, through a sequence of think-aloud modelling, shared writing, guided writing, and independent writing. Students are introduced to the conception that writing is a mode of learning through practices such as writing to interpret raw data. Their motivation to write is developed through practices such as hands-on activities and collaborative work.

Table 1. Content Literacy Framework

Element	Definition and/or Rationale	Illustrative Teaching Practices
Content area literacy with writing focus	Frequent writing contributes to content area learning (e.g., Bangert-Drowns et al., 2004; Boscolo & Mason, 2001); talk supports writing (Nystrand, Gamoran & Carbonaro, 2001).	▪ Writing three or more times per week ▪ Reading to become familiar with analytical genres and learn content ▪ Talk as a prewriting activity ▪ Research using non-text and mixed media, e.g., internet; producing mixed media (e.g., explanations with diagrams)
Conception of writing as learning	Writers' goals guide their planning and revision (Hayes, 1996; Nussbaum, Kardash & Graham, 2005).	▪ Writing to interpret "raw" data and experiences ▪ Discussing the purposes of writing with students ▪ Reflecting on learning following writing sessions ▪ Not providing source materials in the same genre in which students will write to prevent paraphrasing
Education in analytical genres	Argumentation and explanation are central to content area disciplines (Rowan, 1988; Schleppegrell, 2004; Toulmin, 2003).	▪ Reading in analytical genres ▪ Constructing genre knowledge by analyzing models
Analytical genres as heuristics for thinking and learning	Genre guides transformation of information from sources and construction of relationships among ideas in student text (Coe, 1994; Klein, 1999; Spivey, 1991)	▪ Writing arguments to think critically about content ▪ Writing explanations to construct theories
Inquiry writing as preferred lesson type	Writing lessons built around interpretation of concrete data (Graham & Perin, 2007; Hillocks, 2005).	▪ Pre-writing experiences to generate data, e.g., experiments, observations, reading. ▪ Teacher-led lessons on writing strategies ▪ Writing to interpret data.
Cognitive strategy approach to instruction.	Scaffolded instruction in planning and monitoring writing is highly effective (Graham, 2006; Graham & Perin, 2007).	▪ Teaching a strategy for writing each genre ▪ Gradual transfer of responsibility to students: Modeling, shared writing, guided writing, independent writing ▪ Modeling self-monitoring and self-reinforcement
Assessment to scaffold self-assessment, focused on ideas.	Teaching self-evaluation may increase writing achievement (Andrade, Du, & Wang, 2008; Graves, Harris, & Mason, 2005; Olina & Sullivan, 2004).	▪ Instructions to revise focus on genre-based relationships among ideas (e.g., claim-evidence; cause-effect) ▪ Teacher and peer conferencing to scaffold revision ▪ Students using checklists to monitor genre elements in their own writing
Building intrinsic motivation	Positive attitude toward writing predicts variance in achievement (e.g., Bruning & Horn, 2000; Graham, Berninger & Fan, 2007; Guthrie & Cox, 2001).	▪ Interesting topics ▪ Hands-on experiences ▪ Peer collaboration

Assessment is used largely to teach students to self-regulate their writing; for example, students are guided to review their own drafts and mark them up to verify that they have included a variety of genre components. These design principles are conceived as mutually complementary. For example, teaching the conception that writing is a mode of learning complements the use of inquiry as a preferred type of lesson (Hillocks, 2005). Similarly, the decision to teach writing using cognitive strategy instruction complements the practice of using assessment to promote self-regulation with reference to the same strategies. Consequently, in this study, these design elements were introduced as a set, rather than manipulated and tested individually.

As noted, this framework was based on a problem solving model of writing. As with any intervention, it is desirable to understand the mediators and moderators that affect it (Kraemer, Wilson, Fairburn, & Agras, 2002). A *mediator* is a hypothetical causal link that is part of the mechanism through which an intervention operates. By examining mediator variables, it is possible to understand whether an intervention is effective *for the reasons* posited by the model that underpins it. In the model outlined above, genre knowledge and approach to writing are two proposed mediators; that is, instruction is expected to increase genre knowledge and depth of approach to writing, which in turn increase text quality and content understanding. Conversely, a *moderator* is a variable that represents for whom, or under what conditions, an intervention is effective (Kraemer, Wilson, Fairburn, & Agras, & 2002). Moderator variables are important for the practical purpose of determining how to use an intervention most effectively. In the present study, the possible moderators examined were gender and previous writing achievement; more on these below. In the next section, we will focus on the psychological mediators that underpin this framework.

2. Possible Mediators of Content Area Writing and Learning

Genre knowledge. This is one aspect of pragmatic knowledge, a topic that has been a focus for recent discussion in cognitive theories of writing (Alamargot & Chanquoy, 2001). The genres of argumentation and explanation were selected for this project because they play important roles in disciplinary writing and education (Coffin, 2004; Martin, 1989; Schleppegrell, 2004). These are *analytical genres*, that is, texts that foreground relationships among ideas; they fall within the larger category of *informational genres*, that is, texts that convey knowledge. Informational genres appear to have traditionally been under taught in elementary schools (Duke & Bennett-Armistead, 2003; Martin, 1989).

Argumentation is defined here as a type of text, the purpose of which is persuading a reader, or allowing a writer to deliberate on an issue. Argumentation includes components such as a claim, evidence for the claim, warrants that link the claim to the evidence, alternative claims (for the opposing view), rebuttal evidence (for the alternative claim or against the author's claim), countered rebuttals (which attempt to refute the rebuttal evidence), and conclusions (Crammond, 1998; Toulmin, 2003). It

has been proposed that to pursue the rhetorical goal of creating argument moves, writers transform content knowledge (Bereiter & Scardamalia, 1991). For example, a writer may advance a claim, and then search for information to support it. If the writer finds information that supports the claim, then the writer has learned a new claim-evidence relationship; if the evidence refutes the claim, the writer may modify or abandon it, and take up an alternative claim. Extensive previous research has shown that teaching argument strategies comprised of these elements improves students' oral and written argumentation (e.g., Kuhn & Udell, 2003; for a review, see Graham, 2006). However, a point of contention is whether students' explicit knowledge about a text genre *contributes to* writing in that genre (e.g., Freedman, 1993; Johns, 2002). Therefore, a cognitive mediator of interest in this study was *argument genre knowledge*. Our hypotheses were that instruction would increase argument genre knowledge, and that argument genre knowledge would in turn predict unique variance in text quality and learning during writing.

This framework also focused on teaching students to write explanations, which are defined here as texts that tell why or how a phenomenon occurs (Martin, 1989; Schleppegrell, 2004; Veel, 1997). Explanations are central to elementary science and social studies education; for example, students are commonly expected to explain how rocks are formed, how digestion happens, or why a group of states unites to form a nation. Explanation as a genre has been less well researched than argumentation; this limitation is compounded by the fact that some authors refer to explanation using other terms, such as "causal text" (Meyer & Poon, 2001); other authors use the term "explanation" to refer to a broader category of texts that attempt to make ideas intelligible to readers (e.g., Rowan, 1988). Across these differences in terminology and conception, explanations in content area texts are commonly considered to include some subset of the following components: a definition of the phenomenon to be explained; examples; the entities that are involved in the process; the process through which the phenomenon occurs; subexplanations that account for unfamiliar processes that are invoked in the main explanation; and evaluative comments on the significance of the phenomenon (Chambliss, Christenson & Parker, 2003; Macken-Horarik, 2002; Raison, Rivalland, Derewianka et al., 1994; Rowan, 1988).

The model of writing to learn outlined earlier can be applied to explanation writing. The writer initially adopts the rhetorical goal of telling the reader how or why a phenomenon occurs; for example, a student may have the goal of explaining why there are no fish in a local river. She has a text scheme representing the components of explanation, such as causal relationships, examples, and evaluative comments; these are used to generate rhetorical subgoals, such as the elaboration of causal relationships for the reader. To fulfill these rhetorical subgoals, writers set the subgoal of providing the reader with relevant content. To generate this content, they search sources, select relevant information, make inferences from the information, and construct relationships among ideas. In this example, the student searches through a portfolio of information about the river. She finds a graph showing that its pH has fluctuated throughout the

year, and was particularly low in March. To transform this information into knowledge that she can use in an explanation, she must make inferences linking it to the fate of the fish; for example, she could infer that the drop in pH in March is the *cause* of the disappearance. She records this idea in her text, and then continues to build the explanation (Klein & Rose, in review). For examples of explanation texts, see Appendix B.

Textual research indicates that the explanations that students read are initially very concrete, presenting procedures for activities such as how to make a recipe or play a game. Later, explanations begin to address classes of natural and social phenomena, to include multiple causes and effects, and to invoke abstractions (Coffin, 2004; Schleppegrell, 2004; Veel, 1998). Students appear to find explanation writing challenging (Chambliss et al 2003; Schleppegrell, 2004). Moreover, major writing education projects do not teach explanatory writing in the sense described here, although several do teach basic procedural writing (e.g., Culham, 2003; Englert, Raphael, Anderson, Anthony, & Stevens, 1991; Graham & Harris, 2005). Exceptions to this are writing programmes based on systemic functional linguistics, a school which has undertaken extensive analysis of academic texts (Callaghan, Knapp, & Noble, 1993; Macken-Horarik, 2002; Raison et al., 1994; Stead, 2001). Consequently the present study drew on a source informed by systemic functional linguistics, the *First Steps-Writing Resource* (Raison et al., 1994). Based on the theory of writing to learn sketched above, it was hypothesized that explanation genre knowledge would mediate the effects of instruction on dependent variables; that is, instruction would increase explanation genre knowledge; and explanation genre knowledge would predict unique variance in text quality and learning during writing.

Approach to writing. The constructs discussed to this point (text schemata) are cognitive; however, researchers have recognized that attitudes play an important role in writing (Graham, Berninger & Fan, 2007; Hayes, 1996; Hidi, Berndorff, & Ainley, , 2002; Pajares & Valiante, 2001). One attitudinal construct, *approach to writing*, was selected for examination in this study because it refers specifically to an orientation toward making meaning by composing (Lavelle, 1993; 2007). It is conceptualized as two negatively correlated dimensions: *depth*, which is an orientation toward the development of meaning through writing; and *surface*, which is an orientation toward merely completing required writing tasks. This construct was originally assessed using the *Inventory of Processes in College Composition* (Lavelle, 1993), and later the *Inventory of Processes in Secondary Composition* (Lavell, Smith, & O'Ryan, 2002). In an elementary school adaptation of the inventory, items related to elaboration of ideas through writing, and intrinsic motivation, both loaded heavily on the total approaches to writing score (Samuels & Klein, 2008, February).

It was reasoned that writing to learn is an effortful activity; students' choices to elaborate rhetorical subgoals and persist in searching for relevant content knowledge and transforming it, would depend on a deep approach; thus it was a hypothetical

mediating variable of interest. In the instructional framework, it was addressed through teaching practices that introduced students to the conception that writing is a mode of learning, such as asking students to record their ideas before and after a content area writing activity, and then comparing them to note any changes. Their approach to writing was also addressed through activities designed to increase their interest in writing, such as collaboration, hands on experiences, and writing for real audiences (Bruning & Horn, 2000; Hidi & Renninger, 2006). It was hypothesized that instruction in the content area writing framework would increase students' depth of approach to writing; and that in turn, approach to writing would predict unique variance in text quality and learning during writing.

Text quality. Text quality plays two roles in this study. First, from an educational viewpoint, an important purpose of writing instruction, including content area writing instruction, is to teach students to write well. Therefore, text quality appears here as a dependent variable. Second, from a theoretical point of view, the writing process is the mediator of learning; a text reflects the cognitive processes that went into its creation (Sanders & Schilperoord, 2006). Specific features of text that are constituents of quality, such as variety of rhetorical moves and incorporation of content from sources, are also predictors of learning (Klein & Samuels, in press). In this sense, text quality is used here as a proxy for cognitive processes during writing. We hypothesized that text quality would mediate learning during writing; that is, instruction would increase argument and explanation genre knowledge and approach to writing (as outlined above); these would increase text quality; and text quality would predict learning.

3. Possible Moderators of Content Area Writing

A framework for content area writing would ideally benefit all students. A recent meta-analysis of the cognitive strategy instruction literature found that students of various types (e.g., average students, students with learning disabilities) made medium to large gains in writing with this method (Graham, 2006). However, some authors have questioned whether low achieving or novice students can use writing to learn (Ochsner & Fowler, 2004). Classical cognitive theories attribute learning through writing to sophisticated strategies that include subgoal setting and gist revision (Bereiter & Scardamalia, 1987; Flower & Hayes, 1980; see Klein 1999 for a review). Consistent with this, one study has shown that high-achieving students make larger knowledge gains through writing than through talk, while low-achieving students show the reverse pattern (Rivard, 2004). Alternatively, a recent meta-analysis showed a U-shaped developmental curve: Students in Grades 1 to 5 showed evidence of learning from writing intensive curriculum units; those in Grades 6 to 8 did not; and students in secondary and tertiary education again showed such learning (Bangert-Drowns et al., 2004). Because of these previous paradoxical results, it is educationally useful to determine whether this framework is effective for low, medium, and high achieving students. We did not have a specific hypothesis, but simply posed the question of

whether level of previous writing achievement would interact with instruction to affect the posttest variables.

A related issue concerns gender and content area writing education. Most studies have shown that girls outperform boys on writing assessments (Nelson & Van Meter, 2007; Olinghouse, 2008; Salahu-Din, Persky, & Miller, 2008; cf. Graham, Berninger, & Fan, 2007). The reasons for this difference are unclear; however, most girls have a more positive attitude toward writing than most boys (Graham et al., 2007; Hidi et al., 2002). It is possible that traditional school genres such as personal narrative are perceived to be "feminine," and therefore do not appeal to boys (Newkirk, 2000). This view is consistent with the finding that gender orientation, rather than gender itself, is the direct predictor of writing achievement (Pajares & Valiante, 2001). Consequently, some authors have suggested that nonfiction writing may present an opportunity for boys to become more engaged and successful, and an opportunity for both girls and boys to learn to write in non-traditional genres (Newkirk, 2000; Peterson, 2006). Therefore, in the present study, the instruction by gender interaction was examined. In light of previous research, it was hypothesized that girls would score higher than boys on previous writing achievement, but no specific predictions about the instruction by gender interaction were made.

4. The Present Study and Hypotheses

As noted earlier, this content area writing framework was developed in a previous design experiment with one Grade Five instructional class and one Grade Five/Six comparison class (Klein & Rose, in review). Participation significantly increased students' argument genre knowledge, explanation genre knowledge, explanation text quality, and content learning during two science writing activities. It was desirable to replicate this study in a larger quasi-experiment for three reasons. First, in the previous study, the principal author was directly involved in team teaching the content area writing lessons; it was necessary to determine whether this framework would remain effective when classroom teachers implemented it independently. Second, a larger study would allow an examination of moderator variables, including previous writing achievement and gender. Third, a larger study would allow key mediator variables (genre knowledge, approach to writing, text quality) to be examined.

Based on these considerations, the framework was tested in a pretest-posttest quasi-experiment with Grade Five and Six students from nine classes (see Table 2). They initially completed assessments of approach to writing, argument genre knowledge, explanation genre knowledge, and argument text quality. They then participated in approximately six months of content area education based on this framework. Finally, they completed posttests of approach to writing, argument genre knowledge, and explanation genre knowledge. They also completed a sequence of activities designed to assess their ability to use writing as a learning tool: a science prewriting test on nutrition and human organ systems; followed by an argument writing activity; an explanation writing activity; and finally a science posttest on the same topic. The

science posttest score, with science prewriting score accounted for statistically as a covariate, was considered to represent a measure of their ability to learn from writing activities.

Table 2. Research Design

	Instructional Group	**Comparison Group**
October: Pretest assessments	Approach to writing Argument genre knowledge Explanation genre knowledge Argument text quality	
November-April: Instruction	Lessons and writing activities based on content area writing framework	Regular language arts , content area writing
May-June: Posttest Assessments	Approach to writing Argument genre knowledge Explanation genre knowledge Writing to learn assessment (4 phases): Prewriting assessment of science knowledge Argument writing activity (Michael's Snacks) Explanation writing activity: (Gas Exchange) Posttest of science knowledge	

To review, the following hypotheses and questions were examined:

1. *Effects of instruction*: Instruction would significantly increase all writing posttest measures: approach to writing, argument genre knowledge, explanation genre knowledge, argument text quality, explanation text quality, and science knowledge, with corresponding pretest measures accounted for statistically.
2. *Moderators*: (a) Would instruction interact with previous writing achievement to affect each posttest measure? (b) Would instruction interact with gender to affect each posttest measure?
3. *Mediators*: In a path analysis, instruction would contribute to genre knowledge (argument genre knowledge plus explanation genre knowledge) and approach to writing; which would predict text quality (argument text quality plus explanation text quality); which would contribute to science posttest knowledge. See Figure 1 for a summary of this model.

This theory includes three hypotheses:

The *structure strategy hypothesis:* The relationship between instruction and text quality would be mediated by genre knowledge; that is, the path from instruction to text quality would pass through genre knowledge.

The *approach to writing hypothesis*: The relationship between instruction and text quality would be mediated by approach to writing; that is, the path from instruction to text quality would pass through approach to writing.

The *writing to learn hypothesis*: Text quality would predict unique variance in science learning; that is, with prewriting science knowledge accounted for statistically, text quality would account for significant variance in posttest science knowledge.

It may appear somewhat confusing that genre knowledge, approach to writing, and text quality at times are referred to as "mediators" and at times as "dependent variables." However, this is because from a pedagogical viewpoint, text quality, and perhaps approach to writing and genre knowledge, are desirable outcomes of instruction in themselves, making them dependent variables; whereas from a theoretical point of view they are hypothesized to mediate the effect of instruction on the final dependent variable, i.e., learning during writing.

5. Method

5.1 National Context

This study took place in the province of Ontario, in Canada. Canadian educational policy is determined primarily at the provincial level, and implemented in boards of education, which typically comprise a large city or a group of rural counties. Curriculum in content area subjects such as science and social studies is loosely prescribed in guidelines, which recommend topics and objectives for units of study, but do not prescribe the sequence in which these units are taught, or the specific lessons that they include. Curriculum in writing is loosely prescribed in language guidelines, which also prescribe general objectives, such as learning to write for a variety of purposes, and learning to plan and revise, but again, do not prescribe a sequence of specific lessons. The assessment of writing is prescribed through documents that present rubrics and examples of writing at various levels of proficiency. Students complete provincially mandated writing tests in Grade 3 (ages 8 to 9 years) and Grade 6 (ages 11 to12 years), but these are used to assess collective progress rather than determine individual advancement. Genres that are commonly tested include personal narrative, narrative, and personal letter, so these are popular genres for elementary writing instruction. Other genres often taught in elementary schools include procedure (instruction), report, and opinion (argumentation). At the elementary level, students are normally placed into a grade level based on their age; currently, classrooms that combine students from two grade levels are common.

5.2 Participants

All Grade 5 and 6 teachers in two boards of education received an email invitation to participate in the study; nine responded and consented. Subsequently, all students in each class received a letter of information and consent, and 148 students and their parents consented. Eight of the classes could be matched on grade level and proportion of students by school who scored at "expected level" on provincial writing assessments. Four classes were randomly assigned to instruction in the framework, and four to the

comparison condition. Sixteen students did not provide complete data because they relocated to other schools or missed the final assessments for medical reasons. For the experimental analysis, this provided 113 students, including 39 Grade 5 students and 74 Grade 6 students, of whom 58 were girls and 55 were boys. Boards of education in this province place students into grades on the basis of age, so that students at the time of the pretest were 9 yrs, 10 months to 10 years, 9 months of age; students in Grade 7 were 10 years, 10 months to 11 years, 9 months of age. The ninth class was a segregated gifted class, in which 20 students and their parents consented. However, no matching comparison class was available, and this would have skewed results in favor of the instructional group; therefore, the gifted class participated in instruction and assessment, but their data was excluded from the multilevel analysis of the experimental data.

5.3 October Pretest Measures

Several of the following assessments involved ratings, e.g., of written texts. These were carried out by graduate students in educational psychology whose interest was cognitive issues in literacy education. Two were certified elementary school teachers, and one had experience as a tutor at a university writing centre.

Previous writing achievement. This was based on the students' most recent report card writing grade, converted to a numerical score. Teachers in this province assess writing on four dimensions (reasoning, communication, organization, conventions) using a mandated set of rubrics and exemplars, and they periodically have the opportunity to compare their ratings to those of trained assessors. For use as an independent variable (e.g., to test the instruction by previous writing achievement interaction), previous writing achievement was split into three levels: Low, medium, and high. This made the descriptive statistics accompanying the tests for interaction effects more interpretable (e.g., Tables 5 and 7).

Inventory of Processes in Elementary Composition. The purpose of this measure was to assess students' approach to writing. It was adapted from the *Inventory of Processes in College Composition* (Biggs, Lai, Tang, & Lavelle, 1999; Lavelle, 1993). The elementary school version included 40 Likert items comprising an overall deep-surface dimension, in which "deep" comprises the writers' intrinsic motivation to write and intention to elaborate meaning in writing, and "surface" comprises an intention to merely complete required assignments. It has high inter-item reliability, a =.87, and significant predictive validity with respect to writing achievement (Samuels & Klein, 2008, February).

Pretest argument text quality. The purpose of this was to provide an initial assessment of students' analytical writing. Argumentation was chosen because for this genre it is possible to generate prompts that depend on common prior experiences, whereas explanation writing appears to require more specific content knowledge. Students

wrote on the question, "Should students choose the subjects that they study in school?" Two raters independently evaluated the texts holistically on a scale from one to ten. Inter-rater reliability was r = .78. The mean was used to reconcile differences in the ratings.

Argument genre knowledge. The purpose of this survey, and the explanation genre knowledge survey that follows, was to assess students' initial declarative knowledge about analytical genres. To orient students to the argumentation component, it began with the following definition: "A written argument is sometimes called persuasive writing, opinion writing, or exposition. Its purpose is to persuade the reader. One example of a written argument is, 'Why it is Important to Recycle Paper.'" A series of brief, open-ended questions concerning the pragmatic, structural, and lexical aspects of argument followed: Where an argument could be found (1 pt); an example of a title for an argument (1 pt); four kinds of things that could be found in an argument (4 x 1 pt); two words that are clues that a text is an argument (2 x 1 pt); and differences between a good argument and a poor one (2 x 1 pt). The maximum possible score was 10 points. Inter-item reliability based on Guttman's Lamda 5 = .62; inter-rater reliability was r = .86. In a previous study, this survey showed significant predictive validity for argument text quality (Klein & Samuels, in press).

Explanation genre knowledge. For the explanation genre knowledge survey, the introduction read, "An explanation tells *why* or *how* something happens. One example of an explanation is, 'How Airplanes Fly.'" Students answered questions analogous to those in the argument survey: Where an explanation could be found (1 pt); an example of a possible title for an explanation (1 pt); three kinds of things that could be found in an explanation (3 x 1 pt); two words that are clues that a text is an explanation (2 x 1 pt); and differences between a good explanation and a poor one (2 x 1 pt). The maximum possible score was 9 points. Inter-item reliability was Guttman's Lambda 5 = .58; inter-rater reliability was r = .81.

5.4 Professional Development and Instruction

Teachers of the experimental classes participated in professional development sessions approximately every three weeks for the first five months of the study. The first sessions were led by the principal investigator; they included presentations of the theoretical and research background of writing to learn; participation in typical content area writing activities; presentation of teaching strategies; collaborative creation of scripts for possible lessons; and collaborative assessment of writing samples. As the sessions progressed, the teachers led by presenting activities that they had developed, sharing writing samples, and discussing educational problems that arose. The topics of the sessions included: Introduction: Rationale, Framework, Getting Started; The Very Idea of Writing To Learn; Using Forms (Genre) to Promote Learning; A Strategic Approach-- Genre Strategies; A Strategic Approach-- Content Strategies; Strengthening Strategies;

Making Strategies Stick--Self-Regulated Strategy Development (Graham & Harris, 2005); and Review And Conclusion. Important resources for professional development and teaching included the *First Steps: Writing Resource* (Raison et al, 1994), and *Making the Writing Process Work* (Harris & Graham, 1999).

5.5 Instruction

Teachers implemented the framework over a period of approximately six months. The practices that comprised the framework are detailed in Table 1; they have also been described in a previous article (Klein & Rose, in review), and outlined in the introduction above. Each week, teachers were asked to complete at least one session that included teaching about content area writing, and at least one session in which students wrote in a content area subject. These could take place in the same session, for a total of at least 20 lessons. Often, these were organized as a cycle of sessions over the course of approximately one week in which students participated in a hands-on activity or observation; followed by a brief writing lesson; followed by one or more sessions of writing to interpret the activity. Four typical writing lesson sequences are summarized in Table 3.

5.6 Posttest (May) Assessments

Following instruction, several pretest writing measures were repeated as posttests: Approach to writing; argument genre knowledge; and explanation genre knowledge. The posttest assessment of students' ability to use writing as a tool for learning was complex and requires some explanation. Note that its purpose was *not* to assess the content that students had learned during the six months of the instructional phase. Rather, it was a transfer test of their ability to learn from two *new* writing activities. It included four steps: a science prewriting test, an argument writing activity, an explanation writing activity, and a science posttest. The science posttest, with the science prewriting test accounted for statistically, was considered to measure what the students learned during the two writing activities.

Table 3. Four Teacher-Created Lesson Sequences From Instructional Phase Of Project

Lesson Title; Writing Objective	**Conceptual Objective**	**Description of Activity**
***Effects of Contact*: Incorporating evidence into arguments.**	Effects of contact on First Nations people, Europeans.	Students and teacher read arguments, and identified components, including evidence. Students researched selected primary sources on contact between First Nations people and Europeans, particularly contemporary illustrations; students wrote arguments to support their opinion on consequences of contact for both groups.
***Boomerang Lunches*: Writing arguments with counter-arguments and rebuttals.**	Recycling, waste disposal.	Students read, discussed, and evaluated model texts to identify features of good arguments, including counter-arguments and rebuttals. The school was considering boomerang lunches, which aim to reduce disposable packaging of one use food items. Students interviewed members of the school community and researched topic on internet. They wrote arguments for school council, with counter-arguments and rebuttals, to support their position on boomerang lunches.
***Phases of the Moon*: Drafting an explanation**	Mechanism of the lunar cycle.	Students experimented with a light source and sphere to model phases of the moon. Based on this, they wrote an explanation of the lunar cycle. They then exchanged explanations with a peer and wrote a response to the peer's explanation.
***Rube Goldberg machines*: Incorporating scale diagrams into explanations.**	Simple machines can be combined to construct complex machines.	In partners, students tentatively planned Rube Goldberg machines (devices that perform simple tasks in complex, amusing way). They gathered materials, constructed machines, tested them, and modified them. The teacher presented a lesson on adding scale diagrams to explanations. The students then created scale diagrams of their machines, and wrote explanations with references to the diagrams.

Science prewriting test. This was designed to measure students' prior knowledge about human organ systems and nutrition. Categories of questions included: Identifying food groups and classifying foods; evaluating whether various foods are relatively healthy or unhealthy and explaining why; identifying components of the respiratory and circulatory system; and explaining the interaction of the respiratory and circulatory systems. Inter-item reliability was initially low (a = .35); item analysis indicated that two items had low item-total correlations because nearly all students answered them

correctly. When the two low-reliability items were deleted, alpha became .57. Inter-rater reliability on a random stratified sample of 30 tests was very high (r = .94.).

Posttest argument writing task. The purpose was to assess the effect of instruction on the quality of students' argumentation texts and their ability to use argumentation as a means of learning. It was designed to extend the Grade 5 and 6 units of study on nutrition. Pretesting indicated that junior students had a general awareness that a healthy lifestyle included nutritious food, but did not relate this to the constituents of food, e.g., carbohydrates, proteins, and fats; this activity focused on understanding the role of fats in a healthy diet. Each student received a portfolio concerning a fictional student named Michael, which posed the question, "Should Michael's parents make him eat more nutritious snacks?" (Previous experience indicated that students found this more engaging than the direct question, "Should Michael eat more nutritious snacks?"). To provide students with substantive information on which to base their texts, they received a portfolio with relevant source materials. However, these did not present an argument on the issue; to allow students to form their own opinions and compose unique texts, the sources comprised several brief documents in a variety of genres and media, which could plausibly support a positive, negative, or mixed claim. These included: Michael's typical daily menu; his schedule of weekly physical activities; a food guide chart; a height-weight table; "Fast Facts" about the role of fats in a healthy diet; a chart of foods containing various types of fats; and an information sheet on nutrition and disease.

Argument text quality. Students' texts from the *Michael's Snacks* task were holistically rated on a scale from 1 to 10 by two raters from whom the instructional condition and gender of the students were masked; one rater was familiar with the instructional framework and one was not. Raters applied the criterion, "How good is this text as a piece of persuasive writing?" (Holistic rating was used, because later in this research project, a multiple regression analysis will be applied in which holistic text quality will be set as the criterion variable, and the linguistic features of the text will be tested as predictor variables; therefore, it was desirable not to bias this analysis by rating the texts on a specific subset of features). The first rater scored texts on a scale of one to ten and selected a typical text at each level as an anchor. To ensure *intra*-rater reliability, the rater evaluated the texts a second time, reading them from lowest-rated to highest-rated, and verifying that each text that received a given score was better than the texts that received the next lowest score. The second rater, naïve to the original ratings, used the anchor texts as references to assess the remaining texts, and then checked these ratings a second time. Inter-rater reliability was r = .75, $p < .001$; the greatest difference in ratings was 3 points, which occurred for eight texts. These texts were evaluated by a third rater, and in each case, the score was within one point of the mean of the first two ratings. See Appendix A for texts representing a range of scores.

Posttest explanation writing task. The purpose of this task was to assess the effect of instruction on the quality of students' explanation texts and their ability to use explanation writing as a means of learning. Students composed an explanation on the question, "How does the circulatory system exchange gases in the body?" This was intended to extend the regular curriculum, which teaches about the circulatory and respiratory systems, but only in isolation from one another. As with the *Michael's Snacks* task, students received a portfolio comprised of several brief documents to support their writing, including: "Fast facts" about oxygen and carbon dioxide; a CT scan of the lungs showing airways; a diagram of gas exchange in an alveolus; and a schematic of the circulatory system. The source documents did not present an explanation; instead, they provided information that each student could interpret and integrate to create a somewhat unique text.

Explanation text quality. The *Gas Exchange* texts were holistically rated using the same procedure as the *Michael's Snacks* texts. Raters applied the criterion, "How good is this text as a piece of explanatory writing?" Inter-rater reliability was $r = .76$, $p < .001$; eight texts, which differed by 3 points were scored by a third rater, and in each case, the score was within one point of the mean of the first two ratings. Of these eight texts, one came from a participant whose argument text also was also referred to the third rater. See Appendix B for explanations representing a range of quality.

Science posttest. The purpose this task, in conjunction with the prewriting science test, was to assess students' learning during the *Michael's Snacks* and *Gas Exchange* writing activities. It included three sections with different levels of questions; each section included nutrition and organ systems content. The first section comprised cloze items that required recall of information (7 x 1 pt). The second section comprised three comprehension questions requiring students to relate concepts to one another and to think critically (5 pts, 3 pts, 3 pts). The third section comprised two high inference (far transfer) questions that required students to consider a novel scenario, predict what would happen next, and explain why it would happen (2 x 3 pts). Inter-item reliability on standardized items was $a = .62$; inter-rater reliability on a random stratified sample of 30 tests was very high, $r = .96$.

6. Results

6.1 Pretest Assessments

The purpose of the first part of the analysis was to check the comparability of the experimental and comparison groups using the pretest variables, including the combinations of treatment by gender, and treatment by previous writing achievement. Because classes were nested within treatment, a multilevel analysis was used. The dependent variables for the pretest analysis were: previous writing achievement (not trichotomized), pretest approach to writing, pretest argument genre knowledge, pretest

explanation genre knowledge, and pretest argument text quality. To improve interpretability of results, continuous variables were standardized before assessment, so that the parameters approximated standardized coefficients; however, because the independent variables were categorical or have a limited range of values (e.g., gender, treatment, previous achievement level (trichotomized)) their effects on the pretest measures are not truly standardized. As noted above, the eight regular classes, but not the gifted class, were included this analysis (N = 113). Additionally, four participants were removed as outliers based on the Mahalanobis distance, leaving a total of 109 participants.

The results indicated that the instructional group and comparison group did not differ significantly for any pretest measure: Previous writing achievement, F (1, 3.17) = .047, p = .84, β = -.03; pretest argument genre knowledge, F (1, 3.27) = 1.24, p = .34, β = -.32; pretest explanation genre knowledge, F (1, 3.94) = .27, p = .63, β = .13; or pretest argument text quality, F (1, 1.94) = 1.20, p = .39, β = 13. The difference in scores between the two groups on pretest approach to writing approached significance, F (1, 4.38) = 2.61, p = .18, β = -.20. This confirms that the two groups were comparable to one another prior to instruction, except possibly on approach to writing.

Because at posttest we intended to examine the instruction by gender interaction, we tested whether instruction by gender combinations differed significantly with respect to the pretest measures (see Table 4). For conciseness, we will report only on effects that were statistically significant, or approached statistical significance ($p < .20$). In approach to writing, the instruction by gender interaction was statistically significant, F (1, 99.66) = 9.29, p = .003, β= -1.06, such that males in the comparison group scored lower than females in the comparison group, males in the instructional group, and females in the instructional group. Also, on pretest explanation genre knowledge, the instruction by gender interaction was marginally significant, F (1, 99.29) = 3.31, p = .072, β = -.64, such that in the comparison group, girls scored higher than boys, but in the instructional group, girls and boys scored similarly to one another. This suggests that at posttest, for approach to writing and explanation genre knowledge, any instruction by gender interactions would have to be interpreted with caution. (As it turned out, these posttest interactions were not statistically significant). None of the other dependent variables differed with respect to instructional group by gender combinations.

Table 4. Pretest Means (Standard Deviations), By Treatment and Gender

Pretest Variable	Experimental Group		Comparison Group	
	Female (n = 30)	Male (n = 26)	Female (n = 26)	Male (n = 27)
October Pretests				
Previous writing achievement	7.17 (1.97)	6.35 (1.90)	7.04 (1.82)	5.81 (2.30)
Approach to writing	144.13 (19.24)	142.58 (12.31)	147.38 (13.64)	128.00 (16.07)
Argument genre knowledge	3.52 (1.79)	2.67 (1.65)	2.96 (1.49)	2.30 (1.42)
Explanation genre knowledge	3.05 (1.50)	3.13 (1.24)	3.46 (1.24)	2.78 (1.27)
Argument text quality	6.32 (1.21)	5.92 (1.74)	6.79 (1.47)	5.74 (1.26)
Prewriting Assessment				
Science knowledge	10.24 (1.89)	10.34 (1.81)	9.63 (1.82)	9.29 (1.82)

Similarly, we intended to examine instruction by previous writing achievement interactions at posttest, so we compared the corresponding combinations at pretest (see Table 5). Only pretest approach to writing approached a significant effect, F (1, 100.56) = 1.86, p = .16, β = .24, β = .77, such that low achieving students scored higher in the instructional group than the comparison group, medium achieving students scored similarly in both groups, and high achieving students scored higher in the instructional group than the comparison group. Consequently at posttest, an instruction by achievement level interaction in approach to writing would have to be interpreted with caution (although in fact, none occurred).

Prewriting Science Test. Students completed a science test, after the instructional phase and immediately before the writing to learn activities. The experimental group scored slightly higher than the comparison group, but this difference was not statistically significant, F (1, 5.33) = 2.83, p = .15, β = -.45.

6.2 Posttest Measures

Because class was nested within treatment groups, to test the effects of instruction on the posttest measures, multilevel analyses were conducted. For each dependent variable an appropriate pretest covariate was included, e.g., to test of the effects of the independent variables on posttest approach to writing, the covariate was pretest approach to writing. To test Hypothesis 1, that instruction would increase performance on posttest measures, the main effect of instruction was included in the model for each dependent variable. To answer Question 2a concerning whether instruction was effective for students of various achievement levels, the instruction by previous writing achievement interaction was included.

Table 5. Pretest Means, By Treatment and Previous Achievement Level

Pretest Variables	Experimental Group			Comparison Group		
	Low n = 13	Med n = 23	High n = 20	Low n = 17	Med n= 20	High n=16
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean(SD)
October Pretests						
Previous writing achievement	4.15 (1.14)	6.52 (.51)	8.80 (.95)	3.82 (1.13)	6.75 (.44)	8.75 (.93)
Approach to writing	140.38 (15.49)	138.74 (17.96)	150.75 (12.32)	130.00 (17.98)	142.20 (16.62)	139.63 (17.22)
Argument genre knowledge	2.64 (2.05)	2.93 (1.36)	3.65 (1.92)	1.93 (1.37)	2.78 (1.57)	3.23 (1.24)
Explanation genre knowledge	2.13 (1.22)	2.86 (1.13)	3.96 (1.25)	2.59 (1.37)	3.01(1.08)	3.80 (1.21)
Argument text quality	4.92 (1.68)	6.13 (1.16)	6.93 (1.14)	5.50 (1.30)	6.43 (1.14)	6.84 (1.68)
Prewriting Assessment						
Science knowledge	6.48 (1.81)	7.37 (1.18)	7.71 (1.51)	5.46 (1.79)	6.44 (1.02)	6.78 (1.36)

To answer Question 2b, concerning whether instruction is effective for both boys and girls, the instruction by gender interaction was tested. The parameters were partially standardized by converting the dependent variables and covariates to standardized scores prior to analysis; however, the betas cannot be read as fully standardized because the independent variables were categorical or had very few levels (see Tables 6, 7, and 8).

Concerning Hypothesis 1, instruction significantly increased posttest argument genre knowledge, F (1, 5.27) = 12.78, p = .02, β = -.78. It also significantly increased posttest explanation text quality, F (1, 100) = 5.66, p = .02, β = -.39. Instruction had a marginal *negative* effect on science posttest knowledge, F (1, 100) = 3.23, p = .08, β = .47. The effects of instruction on approach to writing, explanation genre knowledge and argument text quality did not approach statistical significance ($p > .20$).

Question 2a asked whether instruction interacted with gender after pretest covariates were accounted for statistically; marginal means are presented in Table 6. This interaction was not statistically significant for any of the dependent variables. However, before concluding that instruction is approximately equally effective for students of both genders, we will report some dependent variables for which this interaction approached statistical significance ($p < .20$). For argument genre knowledge, the marginal means indicate that instruction was slightly more effective for boys than for girls, F (1, 98.73) = 2.57, p = .11, β = -.47. Similarly, for explanation text quality, instruction was marginally more effective for boys than girls, F (1, 100) = 2.17, p = .14, β= -.52. Additionally, it is notable that gender had a *main* effect on posttest argument text quality, such that girls scored significantly higher than boys, F (1, 100), = 5.04, p = .03, β = -.44.

Question 2b asked if instruction interacted with previous writing achievement level, after pretest covariates were accounted for statistically; the marginal means are presented in Table 7. None of the interactions were statistically significant. The interaction effect on argument genre knowledge approached significance, F (2, 99.59) = 1.79, p = .17, β = .68, β = .10, such that instruction was slightly more effective for medium and high achieving students than for low achieving students. Posttest explanation genre knowledge also approached significance, F (2, 100.62) = 1.70, p = .19, β = .30, β =-.52, such that instruction had slightly larger effects on medium achieving students than low or high achieving students. For other posttest assessments (approach to writing, explanation text quality, argument text quality, science knowledge), the instruction by previous writing achievement level interaction did not approach significance.

Table 6. Posttest Marginal Means, By Treatment and Gender

Posttest Variables	Pretest Covariate	Experimental Group		Comparison Group	
		Female (n = 30)	Male (n = 26)	Female (n = 26)	Male (n = 27)
		Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
Approach to writing	**Approach to writing = 140.54**	136.03 (3.19)	136.26 (3.39)	135.71 (3.40)	136.16 (3.57)
Argument genre knowledge	**Argument genre knowledge = 2.89**	5.72 (.39)	6.41 (.41)	4.56 (.41)	4.21 (.41)
Explanation genre knowledge	**Explanation genre knowledge = 3.10**	4.49 (.36)	4.46 (.38)	4.12 (.38)	3.89 (.38)
Argument text quality	**Argument text quality = 6.19**[a]	6.20 (.23)	5.57 (.25)	5.90 (.25)	5.42 (.24)
Explanation text quality	**Previous writing grade = 6.61**	6.09 (.25)	6.22 (.27)	5.86 (.26)	5.23 (.26)
Science knowledge	**Prewriting science knowledge = 6.77**[a]	8.28 (.51)	8.72 (.54)	9.59 (.55)	9.42 (.54)

Note: Superscript a: Pretest task differed from posttest task.

Table 7. Posttest Marginal Means, By Treatment and Previous Writing Achievement

Posttest Variables	Pretest Covariate	Experimental Group			Comparison Group		
		Low n = 13	Med n = 23	High n = 20	Low n = 17	Med n= 20	High n=16
		Mean (SE)	Mean (SD)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
Approach to writing	**Approach to writing=140.54**	129.90 (4.39)	142.86 (3.49)	135.67 (3.73)	134.18 (4.02)	138.09 (3.77)	135.53 (3.95)
Argument genre knowledge	**Argument genre knowledge=2.89**	4.88 (.53)	6.39 (.42)	6.93 (.44)	4.12 (.49)	4.35 (.45)	4.68 (.48)
Explanation genre knowledge	**Explanation genre knowledge=3.10**	3.53 (.49)	4.87 (.40)	5.01 (.41)	3.74 (.45)	3.73 (.43)	4.61 (.44)
Argument text quality	**Argument text quality=6.19[a]**	5.39 (.36)	5.66 (.26)	6.60 (.28)	5.30 (.31)	5.59 (.28)	6.09 (.31)
Explanation text quality	**Previous writing achievement =6.61**	5.88 (.52)	5.93 (.27)	6.66 (.45)	5.13 (.53)	5.54 (.30)	5.97 (.47)
Science knowledge	**Science knowledge=6.77[a]**	9.21 (.75)	7.72 (.57)	8.58 (.63)	9.08 (.72)	9.65 (.61)	9.77 (.68)

Note: Superscript a: Pretest task differed from posttest task.

Table 8. Multilevel Analysis of Posttest Measures

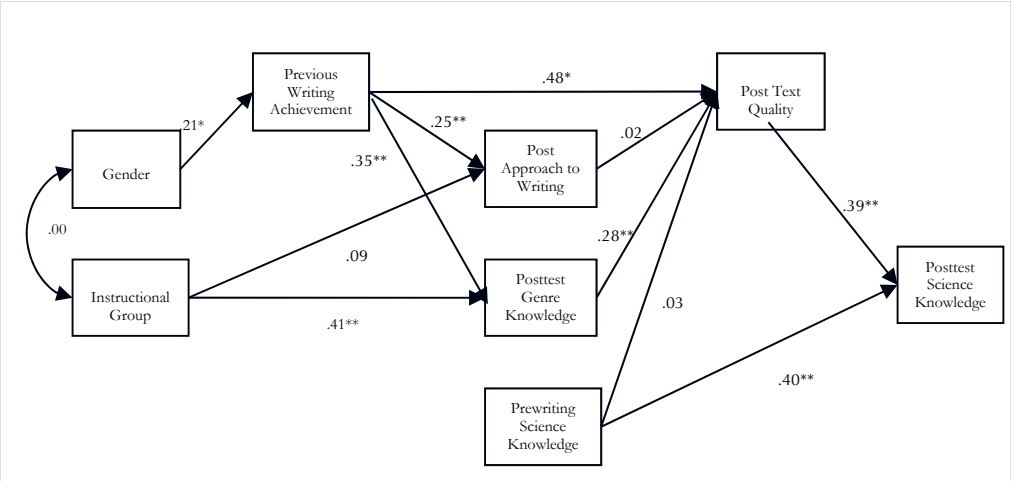
	Posttest Assessments																	
	Approach to Writing			Argument Genre Knowledge			Explanation Genre Knowledge			Argument Text Quality			Explanation Text Quality			Science		
	B	**(SE)**	**p**	**B**	**(SE)**	**p**	**B**	**(SE)**	**p**	**B**	**(SE)**	**p**	**B**	**(SE)**	**p**	**B**	**(SE)**	**p**
Instruction	-.01	(.32)	.97	-.78	(.32)	.02	-.15	(.40)	.70	-.41	(.32)	.21	-.39	(.17)	.03	.47	(.32)	.15
Gender	.01	(.20)	.95	.31	(.21)	.13	-.02	(.25)	.95	-.44	(.23)	.06	.09	(.25)	.72	.14	(.23)	.55
Previous Writing Achievement	-.32	(.28)	.26	-.94	(.27)	.001	-.74	(.36)	.04	-.83	(.33)	.01	-.54	(.58)	.36	.20	(31)	.53
	.39	(.24)	.11	-.24	(.24)	.30	.04	(.30)	.90	-.65	(.27)	.02	-.50	(.36)	.17	-.27	(.26)	.31
Instruction x Gender	.01	(.31)	.97	-.47	(.30)	.11	-.13	(.37)	.73	.11	(.34)	.74	-.52	(.35)	.14	-.19	(.33)	.57
Instruction x Previous Achievement	.24	(.39)	.53	.68	(.38)	.08	.30	(.47)	.54	.28	(.43)	.51	-.04	(.45)	.92	-.42	(.43)	.33
	-.25	(.36)	.48	.10	(.35)	.78	-.51	(.43)	.23	.30	(.39)	.44	.21	(.41)	.61	.23	(.39)	.55
Pretest Covariate	.62	(.09)	.001	.36	(.08)	.001	.13	(.10)	.18	.29	(.09)	.002	.18	(.22)	.42	.58	(.09)	.001
Variance Accounted for by Model	**R^2=49%**			**R^2= 50%**			**R^2=25%**			**R^2=33%**			**R^2=26%**			**R^2=33%**		

It is notable that although previous achievement did not interact significantly with instruction, it showed a strong main effect on several posttest measures, after covariate scores were accounted for statistically (Table 8). This included a significant effect on posttest approach to writing, F (2, 98.53) = 3.18, p = .05, β = -.32, β=.39; argument genre knowledge, F = (2, 98.18) = 4.71, p = .01, β = -.94, β =.24; argument text quality, F (2, 100) = 5.00, p = .009, β= -.83, β = -.65; and a marginal effect on explanation genre knowledge, F (2, 98.63) = 2.67, p = .07, β = -.74, β.= 04. There was no effect on explanation text quality or science knowledge.

6.3 Path Analysis

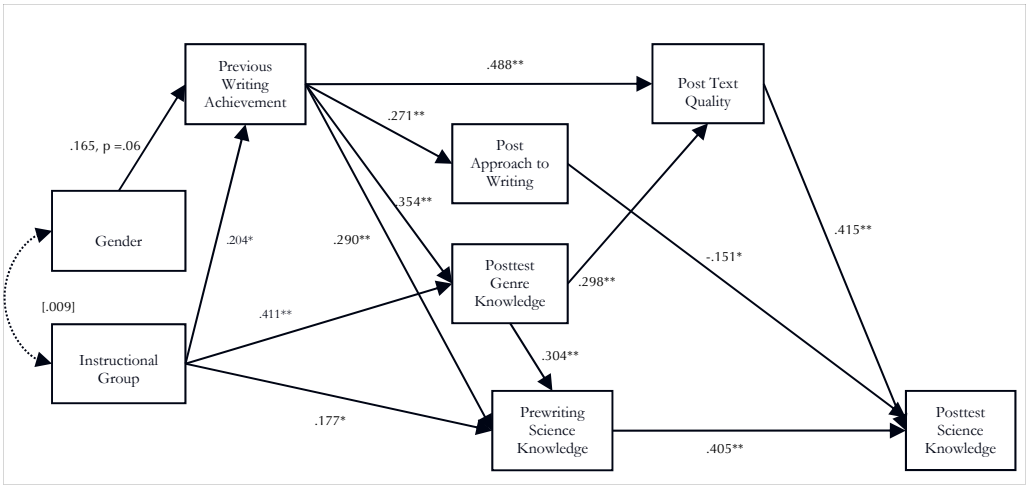
To address Hypotheses 3a, 3b, and 3c, a path analysis was conducted. In a path analysis, the number of variables that can be reliably modelled is constrained by the total number of participants. In the present study, to keep the number of variables appropriate to the sample size, posttest argument genre knowledge and posttest explanation genre knowledge were combined by averaging to create one variable, *posttest genre knowledge*. The inter-item reliability for the combined test was *a* = .71; this was a substantial improvement relative to their separate inter-item reliabilities, which justifies this merger. Similarly, posttest argument text quality and posttest explanation text quality were averaged to create *posttest text quality*; the substantial correlation, r = .56, p < .001 between these two variables supported merging them. Instructional group, gender, and previous writing achievement were included in the analysis because these were important pretest covariates or exogenous variables.

The sample for this analysis included the eight regular classes; additionally, the gifted class was included because orthogonality of predictor variables is not an assumption of path analysis. This brought the total number of participants to 132 students, including 50 students in Grade 5, and 82 in Grade 6; 65 were girls and 67 were boys. The assumptions of path analysis were tested for all continuous variables. As noted earlier, based on the Mahalanobis distance, four outliers were removed from the general classes, and one was removed from the gifted class, bringing the total number of participants to 127. The fact that most gifted students were not eliminated based on the Mahalanobis distance supported the decision to include them in the path analysis. Each variable was normally distributed, except for gender and instruction, which were dichotomous; a scatter plot matrix showed that the relationships among the variables were linear; residual plots confirmed univariate normality. The analysis that followed confirmed that the tolerance for each variable was well above .10 and therefore acceptable.



Note: * < .05; ** < .01

Figure 1. Theoretical Model of Relationships among Variables, With Empirical Values (N = 127).



Note: * p < .05; ** p < .01

Figure 2. Revised Model of Relationships Among Variables (N = 127).

The first path model was created based on hypotheses 3a, 3b, and 3c (see Figure 1). The *structure strategy hypothesis* (3a) was supported by the fact that the path from instruction to posttest text quality was completely mediated by posttest genre knowledge; there was no direct path from instruction to posttest text quality. The *approach to writing hypothesis* (3b), was not supported by the path analysis; the coefficient of the path from posttest approach to writing to posttest text quality was approximately zero. The *writing to learn hypothesis* (3c) was supported by the fact that posttest text quality accounted for significant unique variance in science posttest knowledge, after prewriting science knowledge was accounted for statistically.

As a point of comparison, a second model is presented in Figure 2. Whereas Figure 1 was based on the theory of writing to learn presented in the introduction, Figure 2 was created almost purely empirically. It was constrained only by the consideration that gender and instructional group were exogenous, and posttest science knowledge was the final criterion variable to be explained. The effect of every predictor variable on every endogenous variable was tested, and all relationships significant at the $p = .05$ level were included in the model. One decision was made on a theoretical basis: Posttest genre knowledge and prewriting science knowledge each predicted unique variance in one another; the path direction from posttest genre knowledge to prewriting science knowledge was selected over the reverse path because we reasoned that learning the explanation structure may have facilitated students' science learning during the instructional phase. The empirical model in Figure 2 is largely similar to the hypothetical model. Relative to Figure 1, the three non-significant paths have been trimmed out. Several other paths have been added empirically: a small negative relationship from approach to writing to posttest science knowledge; a path from previous writing grade to prewriting science knowledge; and paths from instructional group to previous writing grade and prewriting science knowledge. The reproduced correlations are presented in Table 9; they fit acceptably for 24 of the 28 observed correlations.

The direct, indirect and total causal effects are presented in Table 10. This model explained 45% of the variance in the ultimate criterion variable, science posttest knowledge; the main determinants of science posttest knowledge were posttest text quality (.42) and prewriting science knowledge (.41). It was also predicted indirectly by previous writing grade (.37). The model also accounted for 45% of the variance in posttest text quality. The strongest determinant of text quality was pretest writing achievement (.60), followed by posttest genre knowledge (.30), and instructional group (.24).

Table 9. Observed and Reproduced Correlations for Revised Path Model

	z1	z2	z3	z4	z5	z6	z7	z8
Observed Correlations (N = 127)								
z1. Gender								
z2. Instructional Group	.01							
z3. Previous Writing Achievement	.17	.21						
z4. Prewriting Science	.01	.38	.46					
z5. Post Approach to Writing	.17	.14	.27	.14				
z6. Post Genre Knowledge	.05	.48	.44	.52	.23			
z7. Post Text Quality	.11	.34	.62	.40	.21	.51		
z8. Post Science	-.08	.19	.36	.55	-.01	.42	.54	
Reproduced Correlations (Revised Model; N = 127)								
z1. Gender								
z2. Instructional Group	.00							
z3. Previous Writing Achievement	.17	.20						
z4. Prewriting Science	.07	.38	.46					
z5. Post Approach to Writing	.05*	.06	.27	.13				
z6. Post Genre Knowledge	.06	.48	.44	.49	.12*			
z7. Post Text Quality	.10	.24*	.62	.36	.17	.51		
z8. Post Science	.02*	.24	.39	.54	-.03	.40	.49	

Note: * = difference between observed and reproduced correlation $\geq$ .10

Table 10. Summary of Causal Effects for Revised Model Shown in Figure 2.

		Model Causal Effects		
Outcome	**Determinant**	**Direct**	**Indirect**	**Total**
Posttest science knowledge	**Gender**	--	.02	.02
(R^2 = .45)	**Previous writing achievement**	--	.37	.37
	Instructional group	--	.24	.24
	Prewriting science	.41**	--	.41
	Posttest approach to writing	-.15*	--	-.15
	Posttest genre knowledge	--	.25	.25
	Posttest text quality	.42**	--	.42
Posttest text quality	**Gender**	--	.10	.10
(R^2 = .45)	**Previous writing achievement**	.49**	.11	.60
	Instructional group	--	.24	.24
	Posttest genre knowledge	.30**	--	.30
Posttest genre knowledge	**Gender**	--	.06	.06
(R^2 = .36)	**Previous writing achievement**	.35**	--	.35
	Instructional group	.41**	.07	.48
Post approach to writing	**Gender**	--	.05	.05
(R^2 = .07)	**Previous writing achievement**	.27**	--	.27
	Instructional group	--	.06	.06
Prewriting science knowledge	**Gender**	--	.06	.06
(R^2 = .36)	**Previous writing achievement**	.29**	.11	.40
	Instructional group	.18*	.20	.38
	Posttest genre knowledge	.30**	--	.30

7. Discussion

The results will be taken up in terms of three questions: Is this content area writing framework effective? Are gender and previous writing achievement significant moderators of the effects of this framework? And, is the model outlined in the introduction, with its mediators, supported by these results?

7.1 Effectiveness of the Framework

We will discuss the positive findings concerning instruction first, followed by the null and negative findings. The most educationally useful result concerned explanation writing. This was the first experimental study, as far as we can determine, that has tested the effect of strategy instruction on students' explanation writing. In contrast, previous strategy instruction research on "explanation writing" has taught procedural writing (instruction), which is a simpler genre than the scientific and social explanations that were the focus here. The need for instruction in explanation writing was supported

by the pretest results documenting students' low level of explanation genre knowledge. The educational value of explanation writing is also underscored by the many elementary science and social studies topics to which teachers applied this genre. Based on these findings, it is recommended that explanation strategy instruction should be included in elementary writing education.

The findings concerning argument writing were rather different. Students in the experimental group, compared to those in the comparison group, did not write significantly better arguments. This finding differed from much previous research, which has shown medium to large effects of strategy instruction on written and oral argumentation (e.g., Kuhn & Udell, 2003; Torrance, Fidalgo & Garćia, 2007; see Graham, 2006 for a review). Discussions with the comparison group teachers suggested an explanation for this: Like the instructional group teachers, they taught and assigned argument writing to students. In contrast, most studies that have produced large effect sizes have used brief, intensive writing interventions, in which the researcher could ensure that the control students did not receive argument instruction during the experiment. However, because the present study took place over a total of eight months, it would have been unethical and impractical to try to prevent comparison teachers from educating students in argument writing.

The other important dependent measure that was not improved by instruction was content learning, as operationalized by science posttest score with prewriting science test score accounted for statistically; in fact, instruction had a marginal negative effect on this variable. This negative effect comprised a pattern in which the prewriting science scores of the instructional group were slightly higher than those of the comparison group, and the comparison group showed slightly greater gains during the writing-to-learn activity than the instructional group. One possible explanation, which treats this difference as real, is that the experimental group was disadvantaged by the fact that during the instructional phase some of their content area education time was invested in writing, rather than other learning activities. However, this is difficult to reconcile with the fact that their prewriting science scores were slightly higher than those of the comparison group. Consequently we prefer a second explanation: This small negative effect could be accounted for by considering the mediating variables represented in the path analysis. As we theorized, instruction increased genre knowledge, which increased text quality, which in turn increased science knowledge. The linkages at each step in this chain are medium in strength: Instruction accounted for approximately 16% of the variance in posttest genre knowledge; which in turn accounted for 9% of the variance in posttest text quality; this in turn accounted for 17 % of the variance in posttest science knowledge. Because the effect of instruction along this path is the *product* of these three relationships, it would be expected to account for less than 1% of the variance in posttest science knowledge. When additional indirect paths are added to the effects of instruction, it contributes 6% of the variance to science posttest knowledge (Table 10). This implied positive effect is very small; given that the observed negative effect is not statistically significant, it could be random.

However, a third explanation, proposed by a reviewer, is also possible: It may be the case the instruction improves genre knowledge, which improves text quality. However, the aspect of text quality improved by genre knowledge may not be the aspect that contributes to learning during writing. That is, genre knowledge may not be the mediator that is associated with writing to learn. In this vein, it is interesting to note that in a previous study on learning through argument writing, most text characteristics that correlated significantly with text quality also correlated with science learning, but there were some exceptions (Klein & Samuels, in press). For example, variety of argument move types in text significantly predicted text quality ($r = .28$, $p < .05$), but not science learning ($r = .12$, n.s.). Consequently, an important question for future research would be to identify what other variables may contribute to learning during writing.

It is informative to compare the results of this study to two other recent studies by the research team. The first was a quasi-experiment on argument instruction; a regression analysis supported the model presented here, but instruction did not affect performance on a posttest of students' ability to use writing to learn (Klein & Samuels, in press). The second study, summarized earlier, was a design experiment that used the same content area writing framework presented here. On a posttest writing activity, it significantly increased science learning; however, it was team-taught by the first author and an experienced elementary teacher with a strong science background. In the present study, the path analysis showed that much of the model of writing to learn was supported, but the effect of instruction on posttest writing to learn was not significant.

Based on these results, it would be too optimistic to claim that this writing framework increases students' ability to use writing as a tool for learning. A useful follow-up study could be a brief intensive intervention that focuses on argument writing, with a control group that does not receive argument instruction. With more intensive strategy instruction, and greater control over the comparison group, the effect on text quality could probably be increased from medium to large (Graham, 2006), which could result in significant effects on content learning. However, the effect of writing instruction on learning would remain indirect, and therefore probably modest in size. Notice that this point should be distinguished from other issues concerning the effectiveness of writing to learn. For example, it is different from the question of whether writing-intensive curriculum units contribute more to learning than non-writing intensive units; they generally do (Bangert-Drowns et al., 2004). This point should also be distinguished from the question of whether scaffolding students in a specific writing activity increases content learning from that activity; research on the Science Writing Heuristic, a cognitively enriched form of laboratory report, indicates that it can (Wallace, Hand, & Prain, 2004).

7.2 Individual Differences in Content Area Writing

The second original contribution of this study was to examine two student characteristics as possible moderators of content area writing instruction. It was of interest that instruction did not significantly interact with previous writing achievement.

Explanation text quality, in particular, showed a significant main effect of instruction, and an inspection of the marginal means indicates that it was approximately equally effective for students at a range of previous writing achievement levels. However, this does not prove that instruction was equally effective across previous achievement levels, particularly for some other posttest measures. In part, this may be because statistical tests for interaction effects are inherently weaker than tests for main effects. It is notable that the interaction of instruction with previous writing achievement level approached significance for two posttest measures: For argument genre knowledge, instruction appears to have been effective for medium and high achieving students, but not for low achievers. Similarly, for explanation genre knowledge, instruction appears to have been somewhat effective for medium-achievers and possibly high achievers, but not for low achievers; however, this last finding is of limited interest, given that the main effect of instruction on this dependent variable was not significant. Therefore, the most reasonable conclusion would be that instruction was effective in improving explanation writing for students at a variety of previous writing achievement levels; but that it may have been effective for improving argument genre knowledge only for medium and high achievers.

The other moderator variable of interest was gender. The finding of principal interest here was that gender did not interact significantly with instruction. However, non-significance cannot be taken to mean that the effects of instruction were equal for students of both genders. There were two interactions that approached statistical significance: Instruction improved argument genre knowledge and explanation text quality, slightly more for boys than for girls. It is of interest that these were the two dependent variables which also showed significant main effects of instruction. Therefore, it might be concluded that for these two variables, boys made gains *at least* equal to girls (cf., Hidi et al., 2002; Olinghouse, 2008). It is notable that girls had significantly higher previous writing achievement. However, the multilevel analysis showed that the effects of gender on all posttest variables, except argument test quality, were nullified by the pretest covariates. Similarly, the path analysis showed that all effects of gender were indirect, passing through previous writing achievement. This implies that although girls started the year as better writers than boys, they appear to have progressed at an approximately similar rate.

It would be premature to suggest that these finding confirm or disconfirm various theories concerning gender and writing. However, they fit well with the analysis by Newkirk (2000), suggesting that boys and girls perform differently in school writing because of the perceived femininity of traditional writing genres such as personal narrative, and that non-traditional, informational genres may allow boys to be more successful than they have been in the past (cf., Pajares & Valiante, 200; Peterson, 2006). To further test this proposal, it would be necessary to juxtapose instruction in personal narrative and analytical genres within the same study. Conversely, the results of the present research are difficult to reconcile with the view that boys have an inherent delay or deficit in language development (Brizendine, 2006; Sax, 2005); if this

were the case, then they would be expected to show smaller gains than girls during instruction. The positive performance of boys in this study is worthy of attention, because stereotypical views about gender and writing affect both students and teachers (Francis, Read, Melling, & Robson, 2003; Pajares & Valiante, 2001; Peterson & Bainbridge, 1999; Peterson & Kennedy, 2006). Methodologically, these findings suggest that rather than proliferating studies that measure pre-existing differences between the writing skills of boys and girls and making educational recommendations based on these, it would be more profitable to examine possible interactions between gender and specific instructional practices (e.g., genre of writing assignments).

7.3 The Theory of Writing to Learn

To the extent that this framework was effective, the question arises of whether this was due to the processes proposed in the problem solving model outlined in the introduction. From this model, three key hypotheses were selected for examination using the mediator variables in the path analysis. The *structure strategy hypothesis* (Meyer & Poon, 2001) was extended to writing by evidence that genre knowledge predicts unique variance in text quality. This extends previous research on students' knowledge about writing, which has demonstrated differences between groups of writers in declarative knowledge about text (e.g., learning disabled students versus regular students), but has not correlated differences in declarative knowledge with differences in text quality, or disentangled the effects of genre knowledge from the effects of other variables such as topic knowledge or previous writing achievement. Additionally, many previous experiments have shown that strategy instruction increases text quality, but research has only begun to examine *why* this is the case (De La Paz, 2007). Our results add to this literature by showing that the effects of instruction on text quality may be mediated entirely by increases in declarative genre knowledge; there was no significant direct path from instruction to text quality bypassing genre knowledge. This weighs against the hypothesis aired by some authors that improvements in writing are based on changes in implicit knowledge rather than explicit knowledge (Freedman, 1993). It is necessary to make a distinction, however, between the finding that explicit knowledge contributes to genre writing, and the claim that explicit *teaching* contributes to genre writing. The latter proposition was recently called into doubt by a large scale study showing that explicitness of instruction did not predict significant variance in students' ability to write reports and procedures, but authenticity of instruction did (Purcell-Gates, Duke, & Martineau, 2007). This is a complex issue; further research should examine the roles of explicit teaching and explicit knowledge, across possible moderator variables such as genre of text, and age of writers.

The present results also were consistent with the *writing to learn hypothesis*; with pretest topic knowledge accounted for statistically, text quality accounted for posttest science knowledge. It can be further noted that the beta weight on the path connecting text quality to posttest science knowledge was very similar to the effect size of writing

on learning found in a recent meta-analysis that compared writing intensive and non-intensive curriculum units (e.g., Bangert-Drowns et al, 2004). A recurring finding then appears to be that writing makes a significant, if typically small, contribution to content learning.

The theory of writing to learn here was based largely on the knowledge transforming model and other problem solving models of writing (Bereiter & Scardamalia, 1987; Flower & Hayes, 1980; Hayes, 1996; Spivey, 1997). Recall that according to this model, (a) writers set rhetorical goals; (b) which evoke rhetorical subgoals; (c) which are transformed into content subgoals; (d) which are solved through content problem solving operations such as inferring. In the present study, no process data was collected, so any inferences about this model are indirect. With this caveat, the fact that the tasks presented rhetorical goals, that is, to persuade and to explain, is consistent with phase (a) of this model. The fact that genre knowledge, which can be considered knowledge about rhetoric, contributed indirectly to learning, is consistent with (b), the rhetorical subgoaling phase of the model. The successful completion of the writing tasks required students to attempt to transform content, as per (c), the content goal setting phase: the argument task required students to attempt to use scientific information to construct relationships such as claim-evidence; and the explanation task required students to attempt to use scientific information to construct cause-effect relationships. These relationships among content elements were not given in the sources, and so they must have been inferred by the students, consistent with (d), the content transforming phase of the model. However, it should be noted that the original knowledge transforming model assumed that the source of content is the writer's long term memory, whereas the present tasks depended heavily on information from sources, as per Spivey (1997). Additionally, the knowledge transforming model assumes a classical cognitive architecture based on formal manipulation of propositions in working memory; other models of writing to learn posit different architectures (e.g., Galbraith, 1999). The data gathered in this study does not discriminate between these architectures.

The element of this model that was not supported was the *approach to writing hypothesis*, which predicted that this variable would mediate the effects of instruction on text quality. It is notable that although there were a range of scores on approach to writing and it had high reliability, it did not contribute significantly to text quality or content learning when other variables were included in the path analysis. More generally, the significant but modest zero-order correlation between approach to writing and posttest text quality (r = .21), and the similar correlation between previous writing achievement and approach to writing, were both nearly identical to correlations in other recent studies linking general attitudinal constructs to writing measures (Graham, Berninger & Fan, 2007; Knudson, 1995). Consistent with this, recent theorizing on approach to learning has proposed that it is dependent on student by environment interactions (Biggs, Lai, Tang, & Lavelle, 1999). This suggests that a modified measure of approach to writing, possibly a generic survey that can be adapted

to whatever specific writing activity a student is about to attempt, would have greater predictive power.

7.4 Limitations

Several limitations have already been acknowledged. To these can be added the issue of treatment integrity. There is an inherent challenge to treatment integrity presented by interventions, such as this one, which invite a high degree of teacher initiative. On one hand, such designs do not allow researchers to make the confident causal inferences that emerge from well-controlled experiments. On the other hand, it is often necessary for teachers to adapt instruction to local conditions (Witt, Noell, La Fleur & Mortenson, 1997). For example, this study was designed to accommodate various classroom teachers who were required to teach different science and social studies curriculum units in different sequences, because of the mix of single grade and multi-grade classrooms unique to each school. More generally, while treatment integrity as traditionally conceived can be sustained over brief, concrete interventions, it tends to decline over time (e.g., Power, Blom-Hoffman, Clarke, Riley-Tillman, Kelleher & Manz, 2005). Consequently, in the present study, we shifted the focus from treatment integrity, to an analogous psychological issue: Did instruction work because it affected the cognitive and motivational mediators that were intended? This consideration is not identical to treatment integrity, because it does not demonstrate that instruction was implemented exactly as recommended. However, this consideration is like treatment integrity in that it focuses on the question of whether instruction worked for the reasons that we hypothesized, here by changing genre knowledge and approach to writing. The path analysis indicates that the answer to this is probably "yes" with respect to genre knowledge, and "no" with respect to approach to writing.

7.5 Conclusion

Analytical genres are central to disciplinary literacy. Elementary students' knowledge of these genres, and their ability to write in them, is limited but can improve with instruction. Moreover, genre knowledge is not simply a means of writing in the correct form; it comprises a resource for generating better texts. The findings from this study point in two directions: On one hand, the markedly indirect relationship between this content area writing framework and content learning implies that it will probably not translate into enduring changes in most students' capacity to use writing to learn. On the other hand, instruction can help students to become better explanation writers. Students who master informational genres possess useful tools for generating good texts and understanding difficult content.

References

Alamargot D., & Chanquoy, L. (2001). *Through the models of writing*. In. G. Rijlaarsdam (Series Ed.), *Studies in writing: Vol. 9*. Dordrecht: Kluwer Academic Publishers. doi: 10.1007/978-94-010-0804-4

Andrade, H. L., Du, Y., & Wang, X., (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice, 27*(2), 3-13 doi: 10.1111/j.1745-3992.2008.00118.x

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. doi: 10.3102/00346543074001029

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

Biggs, J., Lai, P., Tang, C., & Lavelle, E. (1999). Teaching writing to ESL graduate students: A model and an illustration. *British Journal of Educational Psychology, 69*, 293-306. doi: 10.1348/000709999157725

Boscolo, P., & Mason, L. (2001). Writing to learn, writing to transfer. In G. Rijlaarsdam (Series Ed.), & P. Tynjälä, L. Mason, & K. Lonka (Volume Eds.). *Studies in Writing, Volume 7, Writing as a learning tool: Integrating theory and practice* (pp. 83-104). Dordrecht: Kluwer Academic Press.

Brizendine, L. (2006). *The female brain*. New York: Morgan Road Books.

Bruning, R., & Horn, C. (2000). Developing motivation to write. *Educational Psychology, 35*, 25-37. doi: 10.1207/S15326985EP3501_4

Callaghan, M., Knapp, P., & Noble, G. (1993). Genre in practice. In B. Cope, & M. Kalantzis (Eds.), *The Powers of literacy: A genre approach to teaching writing* (pp. 179-202). London: Taylor & Francis.

Chambliss, M. J., Christenson, L. A., & Parker, C. (2003). Fourth graders composing explanations about the effects of pollutants: Writing to understand. *Written Communication, 20*, 426-454. doi: 10.1177/0741088303260504

Coffin, C. (2004). Learning to write history: The role of causality. *Written Communication, 21*, 261-289. doi: 10.1177/0741088304265476

Crammond, J. G. (1998). The uses and complexity of argument structures in expert and student persuasive writing. *Written Communication, 15*, 230-268. doi: 10.1177/0741088398015002004

Culham, R. (2003). *6 +1 traits of writing: The complete guide (grades 3 and up)*. New York: Scholastic Teaching Resources.

De La Paz, S. (2007). Managing the cognitive demands for writing: Comparing the effects of instructional components in strategy instruction. *Reading & Writing Quarterly, 23*. 249-266. doi: 10.1080/10573560701277609

Duke, N. K., & Bennett-Armistead, S. (2003). *Reading and writing informational text in the primary grades: Research-based practices*. New York: Scholastic.

Englert, C. S., Rapahel, T. E., Anderson, L. M., Anthony, H. M., & Stevens, D. D. (1991). Making strategies and self-talk visible: Writing instruction in regular and special education classrooms. *American Educational Research Journal, 28*, 337-372. doi: 10.3102/00028312028002337

Flower, L., & Hayes, J. (1980). The cognition of discovery. *College Composition and Communication, 31*, 21-32. doi: 10.2307/356630

Francis, B., Read, B., Melling, L., & Robson, J. (2003). University lecturers' perceptions of gender and undergraduate writing. *British Journal of Sociology of Education, 24*, 357-373. doi: 10.1080/01425690301891

Freedman, A. (1993). Show and tell? The role of explicit teaching in the learning of new genres. *Research in the teaching of English, 27*, 222-251.

Galbraith, D. (1999). Writing as a knowledge-constituting process. In M. Torrance & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 139-159). Amsterdam: Amsterdam University Press.

Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187-207). New York: The Guildford Press.

Graham, S., Berninger, V., & Fan, W. (2007). The structural relationship between writing attitude and writing achievement in first and third grade students. *Contemporary Educational Psychology, 32*, 516-536. doi: 10.1016/j.cedpsych.2007.01.002

Graham, S. & Harris, K. (2005). *Writing better: Effective strategies for teaching students with learning difficulties*. Baltimore, MD: Paul H. Brookes Publishing Company.

Harris, K, & Graham, S. (1999). *Making the writing process work*. Brookline, MA: Brookline Books.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-27). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Hidi, S., Berndorff, D., & Ainley, M. (2002). Children's argument writing, interest and self-efficacy: an intervention study. *Learning and Instruction, 12*, 429-446. doi: 10.1016/S0959-4752(01)00009-3

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*, 111-127. doi: 10.1207/s15326985ep4102_4

Hildebrand, G. M. (1998). Disrupting hegemonic writing practices in school science: contesting the right way to write. *Journal of Research in Science Teaching, 35*, 345-362. doi: 10.1002/(SICI)1098-2736(199804)35:4<345::AID-TEA4>3.0.CO;2-Q

Hillocks, G. (2005). At last: The focus on form vs. content in teaching writing. *Research in the Teaching of English, 40*, 238-248.

Johns, A. M. (2002). Introduction: genre in the classroom. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 3-13). Mahwah, NJ: Lawrence Erlbaum Associates.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394. doi: 10.1037/0033-295X.85.5.363

Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review, 11*, 203-270. doi: 10.1023/A:1021913217147

Klein, P. D., Boman, J. S., & Prince, M. P. (2007). Developmental trends in a writing to learn task. In M. Torrance, D. Galbraith, & L. Van Waes (Eds.), *Writing and Cognition: Research and Application* (pp. 201-217). Amsterdam: Elsevier.

Klein, P. D., & Rose, M. A. (in review). Learning to write to learn.

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

Knudson, R. (1995). Writing experiences, attitudes, and achievement of first to sixth graders. *Journal of Educational Research, 89*, 90-97. doi: 10.1080/00220671.1995.9941199

Kraemer, H. C., Wilson, T., Fairburn, C. G., & Agras, S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59*, 877-883. doi: 10.1001/archpsyc.59.10.877

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*, 1245-1260. doi: 10.1111/1467-8624.00605

Lavelle, E. (2007). Approaches to writing. In M. Torrance, L. V. Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 219-230). Elsevier: Amsterdam. doi: 10.1108/S1572-6304(2007)0000020015

Lavelle, E., Smith, J., & O'Ryan, L. (2002). The writing approaches of secondary students. *British Journal of Educational Psychology, 72,* 399-418. doi: 10.1348/000709902320634564

Leonard, S. (2005). *Why gender matters: What parents and teachers need to know about the emerging science of sex differences*. New York: Doubleday.

Macken-Horarik, M. (2002). "Something to shoot for": A systemic functional approach to teaching genre in secondary school science. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 17-42). Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, J. R. (1989). *Factual writing: Exploring and challenging social* reality (2nd ed.). Oxford, UK: Oxford University Press.

Meyer, B. J. F., & Poon, L. W., (2001). Effects of structure strategy training and signaling on recall of text. *Journal of Educational Psychology, 93*, 141-159. doi: 10.1037/0022-0663.93.1.141

Moss, B. (2005). Making a case and a place for effective content literacy instruction in the elementary grades. *The Reading Teacher, 59*(1), 46-56. doi: 10.1598/RT.59.1.5

Nelson, N. W., & Van Meter, A. M. (2007). Measuring written language ability in narrative samples. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 23,* 287-309. doi: 10.1080/10573560701277807

Newkirk, T. (2000). Misreading masculinity: Speculations on the great gender gap in writing. *Language Arts, 77*, 294-300.

Ochsner, R., & Fowler, J. (2004). Playing devil's advocate: Evaluating the literature of the WAC/WID movement. *Review of Educational Research, 74*, 117-140. doi: 10.3102/00346543074002117

Olinghouse, N. G. (2008). Student and instruction level predictors of narrative writing in third-grade students. *Reading and Writing, 21*, 3-26. doi: 10.1007/s11145-007-9062-1

Pajares, F., & Valiante, G. (2001). Gender differences in writing motivation and achievement of middle school students: A function of gender orientation? *Contemporary Educational Psychology, 26*, 366-381. doi: 10.1006/ceps.2000.1069

Peterson, S. (2006). Influence of gender on writing development. In MacArthur, C. A., Graham, S., & Fitzgerald, J. (Eds.), *Handbook of Writing Research* (pp. 311-323). New York, N.Y.: Guilford Press.

Peterson, S. & Bainbridge, J. (1999). Teachers' gendered expectations and their evaluation of student writing. *Reading Research and Instruction, 38*, 255-271. doi: 10.1080/19388079909558293

Peterson, S. S., & Kennedy, K. (2006). Sixth-grade teachers' written comments on student writing: Genre and gender influences. *Written Communication, 23*, 36-62. doi: 10.1177/0741088305282762

Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. H. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools, 42*, 495-507. doi: 10.1002/pits.20087

Purcell-Gates, V., Duke, N. K., & Martineau, J. A. (2007). Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching. *Reading Research Quarterly, 42*, 8-45. doi: 10.1598/RRQ.42.1.1

Raison, G., & Rivalland, J. (1994). *First Steps: Writing Developmental Continuum*. Porstmouth, NH: Heinemann.

Raison, G., Rivalland, J., Derewianka, B., Johnson, T. D., Sloan, P., Latham, R., Kovalevs, K., Bindon, R., Annandale, K., Larsen, J., & Dewsbury, A. (1994). *First Steps: Writing Resource*. Portsmouth, NH: Heinemann.

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

Rowan, K. E. (1988). A contemporary theory of explanatory writing. *Written Communication, 5*, 23-56. doi: 10.1177/0741088388005001002

Salahu-Din, D., Persky, H., & Miller, J. (2008). *The Nation's Report Card: Writing 2007* (NCES 2008–468). Washington, D.C: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Samuels & Klein (2008, February). Approaches to writing in elementary students. Paper presented at the Third Annual Santa Barbara Conference on Writing Research: Writing Across Borders, Santa Barbara, California.

Scardamalia, M., & Bereiter, C. (1991). Literate expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 172-194). Cambridge: Cambridge University Press.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Sanders, T. J. M. & Schilperoord, J. (2006). Text structure as a window on the cognition of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 386-402). New York: The Guildford Press.

Spivey, N. N. (1997). *The constructivist metaphor: Writing, reading, and the making of meaning*. San Diego, CA: Academic Press.

Stead, T. (2001). *Is that a fact? Teaching nonfiction writing K-3*. Portland, ME: Stenhouse Publishers.

Torrance, M., Fidalgo, R., Garćia, J-N. (2007). The teachability and effectiveness of cognitive self-regulation in sixth-grade writers. *Learning and Instruction, 17*, 265-285. doi: 10.1016/j.learninstruc.2007.02.003

Torrance, M., Thomas, G. V., & Robinson, E. J. (1996). Finding something to write about: Strategic and automatic processes in idea generation. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 189-206). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Toulmin, S. (2003). *The uses of argument* (updated ed.). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511840005

Tynjälä, P. (2001). Writing, learning and the development of expertise in higher education. In G. Rijlaarsdam (Series Ed.), & P. Tynjälä, L. Mason, & K. Lonka (Volume Eds.). *Studies in Writing, Volume 7, Writing as a learning tool: Integrating theory and practice* (pp. 37-56). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Veel, R. (1997). Learning how to mean—scientifically speaking: Apprenticeship into scientific discourse in the secondary school. In F. Christie & J. R. Martin (Eds.), *Genre and institutions: Social processes in workplace and school* (pp. 161-195). London, UK: Continuum International Publishing Group.

Wallace, C. S., Hand, B., & Prain, V. (2004). *Writing and learning in the science classroom*. Dordrecht: Kluwer. doi: 10.1007/978-1-4020-2018-6

Witt, J. C., Noell, G. H., La Fleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30*, 693-696. doi: 10.1901/jaba.1997.30-693

Author Note

This project was funded by a grant from the Social Sciences and Humanities Research Council of Canada. We wish to thank Boba M. Samuels, Heather Brown, and Jennifer Sinclair for their assistance in collecting and analyzing data for this study. We also thank the teachers and students of the Thames Valley District School Board, and the London District Catholic School Board for their participation in this project. Finally, we would like to thanks the editors and reviewers for their advice, and particularly thank Dr. Huub van den Bergh, for his helpful advice on the statistical analysis.

Appendix A: Posttest Argument Samples

Should Michael's Parents Make Him Change His Snacks?

With Text Quality Score (Original spelling and punctuation)

Argument #131, Text quality = 4.0

Should Michael's Parent's Make Him Chang His Snacks

I think Michael is a healthy kid but I think he eats a little to much for brekfast and lunch. If he changes some of the foods then he is pretty healthy.

If Michael would cut back on some of the stuff that he eats for breakfast like just having some cold cereal an apple and some milk that would be good. Just one little thing can make a difference.

I think potato chips isn't the best morning snack maby if he could have an orange for morning snack and potato chips for an evning snack or even an after noon snack.

For dinner I don't think he need ice cream almost every day, maby he could have it once a week insted.

So Micheal just needs to cut back on some things like with dinner and breakfast.

Argument # 111, Text quality = 6.5

I think that Micheal's parents shouldn't change his snacks because if his parents give him something the Micheal hates, then Micheal will stop eating. Micheal is a very healthy boy, he does lots of exersise all year round. If Micheal is doing really well in school, like he works hard, gets good marks, and does his homework, then the foods that he eats are helping him do well at school. Micheal is not overweight, even though he eats chips and fries, but even though he eats junk food, he still eats healthy foods, like green beans, green peppers, and a healthy salad. Micheal eat very well balenced meals. He eats grain, protein, dairy, and a little bit of junkfood. But a little bit of junkfood isn't that bad, he needs some junkfood to help him with sports. Micheal also is average height which means that he's eating enough foods.

Argument #40, Text quality = 9.0

Should Michael's Parents Change His Snacks?

Yes, the usual snacks that Michael eats are unhealthy (doughnuts, chips, and almonds) except for almonds. Although Michael is a phisically fit boy (Height - 141 cm Weight - 34 kg) as he eats more junk food he will gain more and more weight until he is obese. He may exercise everyday he also needs to eat well too or he will still gain weight.

First of all, eating a doughnut can give you a lot of fat, same thing with potatoe chips. Potatoe chips may be made out of potatoes, but when they are fried to turn into the actual potatoe chip it loses all it's nutrients. It also absorbs grease when it is being fried and gets more fat. A doughnut also contains a lot of sugar which can rot your teeth.

Second of all, eating foods that contain a lot of fat can cause stroke or a heart attack which can lead to death. Stroke and heart attacks happen when fat blocks the arteries and doesn't let blood go threw. This causes damage to some parts of your body.

Some people say that when Michael eats healthy foods at breakfast, lunch, or dinner the fat gets stoped. Well the truth is, it doesn't.

So Michael should really change his snacks, why would he want to eat those disgusting things anyways? I think Michael should eats nuts, seeds, and olives instead, they're a lot better than doughnuts or potatoe chips.

Appendix B : Posttest Explanation Samples:

How Do Our Bodies Exchange Gases?
With Text Quality Score (Original spelling and punctuation)

Explanation # 93, Text quality = 4.0

Our bodies need to exchange gas and we can't just do it ourselfes. We need help from our lungs and our alveoli. The alveoli has blood vessels wraped around it. This is were gasses move between blood. You inhale and exhale through the Alvieoli. The also lock oxygen in and cardon dioxide.

Explanation #8, Text quality score = 6.5

It all strarts by breathing in oxgen. Inside of the lungs are thousands of little air sacs calles Alveoli. Each Alveoli has blood vesles wrapped around it. This is where gases move between the lungs and the blood. The Artaries take blood with oxygen to all the tissuse. The the vienes carry blood out of the tissuse and all the way back to the heart and lungs. We the breath out and out comes the carbon dioxide.

Explanation #41, Text quality score = 9.0

How do our Bodies Exchange Gases?
Our muscles need oxygen to survive. They get the oxygen from the blood. The blood get it from the lungs. The lungs get it from the nose/air vessel. We breath in oxygen and exhale carbon dioxide.

<u>What are lungs?</u>
The lungs are a part of the bodie that have thousands of tiny air sacs called alveoli. The alveoli take the air and give it to the blood. Also they take the carbon dioxide out of the blood and into the air vessel.

<u>How does this happen?</u>
As you see on the diagram on the right, oxygen goes into the blood and carbon dioxide comes out. The blood with the oxygen flows into the left side of the heart and through to the muscles. The muscles take the oxygen and fill the blood with carbon dioxide. That blood then flows through the right side of the heart into the lungs. There the alveoli exchange the gases and the cycle repeats. This happens in a matter of seconds.

[diagram of alveoli]
[diagram of circulatory system]

Appendix C: Scoring Argument Genre Knowledge

1. Where could you find a written argument? (max. 1 point.)

1 pt.: Any place where written arguments are probable, e.g. editorials, letter to the editor.

0.5 pts: Vague but valid responses, e.g., newspapers, books, the internet, magazines.

2. Give an example of a title for a written argument. (max. 1 pt.)

0.5 pts: Title refers to a topic that is potentially controversial, e.g., taxes, National Hockey League lockout, election, school uniforms.

0.5 pts: Title refers to the topic in a way that implies it is controversial, e.g. "Are taxes too high?" "Taxes are too high," or "Cut taxes!"

Total of 1 mark for a controversial topic and a title that reflects that it is controversial.

3. What are four important kinds of things to include in a written argument? **(**max. 4 pts.**)** Points should be specific to argumentation. 1 pt. for each different element of an argument. Students may refer to these elements using unconventional terms.

thesis / opinion / claim /statement

introduction / importance / background

reasons / arguments / proof / "why" / support /examples / facts / benefits

arguments against / other opinion / opposing view / "why not" / negatives for your side / positives for the other side

rebuttal / negatives for the other side

conclusion / summary

Students receive 0.5 pts each for other valid features not specific to argumentation: title, details, main idea, "beginning, middle, end," to a maximum of 2 points.

4. List two words that are clues that something is a written argument (max. 2 pts)

1.0 pts per term: Must be terms that can express logical relations associated with arguments, e.g., therefore, because, so, why, but, however, agree, disagree, although, right, however, opinion, think, reason for, reason against, research states, yet, should.

5. What do you think makes the difference between a good written argument and a poor one? (max. 2 pts)

Responses may not duplicate answers to earlier questions, e.g., to Question 4.

0 pts: No response, wrong answer, not at all specific to argumentation, e.g., "good spelling."

0.5 points: Obvious, nearly tautologous, e.g., "A good argument persuades the reader"; vague, e.g., "A good argument is well written."

1 point: Lists 1 distinction between good and poor arguments that is relevant to the logical function of argumentation (shows that the claim is true) and/or the rhetorical function of argument (persuades the reader that claim is true); may be relatively easy "A good argument gives reasons;" may be only partially specific to argumentation, e.g. "gives details" "on topic" "interesting" "writer seems enthusiastic."

2 points: Lists 2 or more points as noted above; or chooses 1 point and develops it well, must be non-obvious, e.g., "well researched," "considers both sides of an issue," "has more reasons for the opinion," "proves point and gives reasons against opposing point."

Maximum Total: 10 points.

Appendix D: Scoring Explanation Genre Knowledge

1. Where could you find an explanation? (1 point)

1.pt: Any place where written explanations are probable, e.g., textbooks, science books, "How Things Work" internet site.

0.5 pts: Vague answers, e.g., books, the internet, magazine

2. Give an example of a possible title for an explanation (1 point)

0.5 pts: Title refers to a topic that invites causes and effect explanation, e.g., tides, electric motors, World War 1.

0.5 pts: Title frames the topic in a way that indicates that it is an explanation. e.g., "Why…, how…, the operation of…, the causes of… the effects of…"

1.0 pts: Both the topic and the frame are appropriate, e.g., How electrical motors work; why there are tides, what caused World War 1

3. What are three important kinds of things to include in an explanation? (max 3 points total)

Points should be specific to explanation. 1 pt. for each different element of an explanation. Students may refer to these elements using unconventional terms.

Definition / topic / identifying topic / what it is about

Components / parts / things / elements / objects

Operations / what happens / how it works/ why it works / steps / sequence of events

Applications / where it is used / other examples

Evaluation /why it is important / interesting comments / conclusion / special features

Diagram

Students receive 0.5 pts each for other valid features not specific to explanation, e.g., Title, details, main idea, "beginning, middle, end," to a maximum of 2 points.

4. List two words that are clues that something is an explanation (max 2 points total)

1.0 pts per term: Must be terms that can express logical relations associated with explanations; should suggest cause-and-effect relations, how or why things happen, e.g., "causes, because, makes, affects"

5. What do you think makes the difference between a good explanation and a poor one? (max 2 points).

0 pts: No answer; wrong answer; not relevant to compositional issues, e.g., "has good grammar," "good spelling."

0.5 pts: Obvious, nearly tautologous, e.g., "A good explanation tells why something happens," "A good explanation is well written."

1 point: Student makes 1 distinction between good and poor explanations that is relevant to the logical function of explanation, e.g., "tells all the steps," "steps are in the right order," "tells why each event happens," "includes all the reasons that something happens," "includes a diagram."

May be only partially specific to explanation, e.g., "gives details" "on topic" "interesting" "writer seems enthusiastic," but can receive a maximum of 1 point in total for 1 or more responses of this kind.

2.0 pts: Lists 2 or more points similar to those above, or chooses 1 point and develops it well.