

Supporting Elementary English Language Learners' Argumentative Writing
Through a Functional Grammar Approach

by

Catherine Lynne O'Hallaron

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Educational Studies)
in the University of Michigan
2014

Doctoral Committee:

Professor Mary J. Schleppegrell, Co-Chair
Professor Annemarie Sullivan Palincsar, Co-Chair
Professor Nell K. Duke
Professor Emerita Diane Larsen-Freeman

© Catherine L. O'Hallaron 2014

DEDICATION

A mi hija Rocío, el gran tesoro de mi vida. Tu mamá te ama, ahora y para siempre.

ACKNOWLEDGMENTS

As years of thinking, talking, and writing about this project finally come to fruition, I would like to express my sincere appreciation for those who made it possible in the first place. My dissertation co-chairs, Mary Schleppegrell and Annemarie Palincsar, designed the incredible research project from which my study was born, and I consider myself enormously fortunate to have been able to play a role in that work and soak up their combined wisdom. Their guidance and encouragement regarding my own work were invaluable from beginning to end. As my academic advisor, Mary Schleppegrell has been a model of the kind of researcher and mentor that I hope to be. Her ongoing role in sharpening my thinking, providing new professional experiences, and encouraging me to pursue challenging opportunities has helped get me to where I am today, and her celebration of her students in all aspects of their lives allowed me to bring my whole self to graduate school, which enriched my experience immeasurably. I would also like to thank committee members Nell Duke and Diane Larsen-Freeman for taking the time to read my work and sharing their reactions; their insightful feedback strengthened the presentation of my research and pushed me to think about it in novel ways, a delightful experience after having been so close to my data for so long.

On a larger scale, this project was made possible by the support and inspiration I have received not just over the last few years, but throughout my academic life, from a league of extraordinary women. In elementary and middle school, Janice Schade and Beth Lockwood made special efforts—likely long forgotten to them, but never to me—to

nourish my mind and spirit through books that sit on my shelf to this day. Pamela Ashmore and Sheilah Clarke-Ekong, undergraduate professors turned mentors and friends, are beloved to me always for their many years of wise words, listening ears, and unwavering belief in my potential. Many thanks to Norine Berenz for feeding my linguistics addiction during my MA and helping inspire me to study further.

And to my friends and family, a thousand thank-yous more. B.J. O'Hallaron, Lori O'Hallaron, Erin O'Hallaron Davis, Connie O'Hallaron, Laurel Zmolek-Smith, Veronica Vitté, Dionisia "Negra" Velázquez, and Ezequiel Bayuelo, your friendship has sustained me even as we've scattered across the country and the globe. Dan O'Hallaron, the genes didn't hurt. Jason Moore, Shannon Schmoll, Emily Mihocko, and Jean Mrachko, thank you for riotous laughter, heartfelt commiseration, and for being the people in whose company I found the communion I came to grad school to seek. Victor Brian Montaña Martínez has consistently been as supportive a husband as anyone could ask for, and has the frequent flyer miles to prove it. Gracias, amor. Rocío Marina Montaña O'Hallaron, escribí esta tesis por y para ti. You were instrumental in keeping me productive and in bringing true joy to a long and sometimes tedious process. Thank you, mi linda.

Finally, I would like to thank the University of Michigan School of Education, the Horace H. Rackham School of Graduate Studies, the National Academy of Education, and the Spencer Foundation for their financial support, without which I would not have been able to complete this work.

TABLE OF CONTENTS

Dedication	ii
Acknowledgments	iii
List of Tables	vi
List of Figures	viii
List of Appendices	ix
Abstract	x
Chapter 1: Introduction	1
Chapter 2: Review of Literature	11
Chapter 3: Methods	46
Chapter 4: Second-Grade Writing	68
Chapter 5: Fourth-Grade Writing	125
Chapter 6: Discussion	214
Appendices	262
References	298

List of Tables

2.1 Metalanguage used to support argumentative writing	45
3.1 Overview of research questions, data sources, and analysis strategies	49
3.2 Second-grade focal students	53
3.3 Fourth-grade focal students	53
3.4 Approach to clause-level analysis	57
4.1 Inter-rater agreement (exact and adjacent) for second-grade pre- and post-tests	82
4.2 Average pre- to post-test changes in scores by subgroup, 2 nd grade	85
4.3 Criteria used to evaluate Evidence and Reason	100
4.4 Criteria used to evaluate Counterargument and Evaluation	106
4.5 Evaluative framework for argument (second-grade version)	111
4.6 Average scores given by independent raters to student subgroups	121
5.1 Inter-rater agreement (exact and adjacent) for fourth-grade pre- and post-tests	151
5.2 Average pre- to post-test changes in scores by subgroup, 4 th grade	154
5.3 Criteria used to evaluate Description of Problem and Claim	171
5.4 Criteria used to evaluate Restatement of Claim	174
5.5 Criteria used to evaluate Evidence and Reason	177
5.6 Criteria used to evaluate Counterargument and Evaluation	181
5.7 Evaluative framework for argument (fourth-grade version)	189
5.8 Average scores given by independent raters by claim	208

5.9 Average scores given by independent raters to by stage	210
5.10 Average scores given by independent raters to by subgroup	212
6.1 Clause-level descriptive statistics for second- and fourth-grade writing	217
6.2 Grade level differences in performance	221
6.3 Examples of strategies for Evaluation used by fourth graders	233
6.4 Pedagogical recommendations by grade level	236
6.5 Comparison of average change in scores by subgroup, 2 nd and 4 th grades	244
6.6 Evaluative framework for argument (fourth-grade version)	248

List of Figures

3.1 Michigan's English Language Proficiency Assessment (ELPA) levels	52
3.2 The teaching-learning cycle	63
4.1 Samir's handwritten argument	114
4.2 Hassan's handwritten argument	119
5.1 Isa's handwritten argument	195
5.2 Sanaa's handwritten argument	199
5.3 Amir's handwritten argument	204

List of Appendices

Appendix A: Focal student interview protocol	262
Appendix B: Scoring materials for independent raters	263
Appendix C: Graphic organizers for planning and writing, second grade	268
Appendix D: Graphic organizers for planning and writing, fourth grade	272
Appendix E: Bald Eagles text	277
Appendix F: Model arguments for second grade	278
Appendix G: Grizzly Bear fact sheet	279
Appendix H: Cane Toads text	280
Appendix I: Lungworms text	282
Appendix J: Model arguments for fourth grade	283
Appendix K: Text type overview handout	284
Appendix L: Meat Ants text	285
Appendix M: Stage analysis coding manual	286
Appendix N: Percentages of exact and adjacent agreement for primary trait scoring	292
Appendix O: Independent Scores, Individual Results	293

ABSTRACT

Academic standards recently adopted by a majority of U.S. states promote argumentative writing as a cornerstone of K-12 education, to be taught from the earliest grades on. This creates a pressing need for effective instructional approaches for teaching argumentation in elementary schools. Empirical evidence of the argumentative abilities of very young students, however, is sparse, with a particular gap in research on argumentation in young English language learners (ELLs). To accurately speak to children's potential to develop effective arguments, we need to know what kind of writing they produce in appropriately supportive instructional contexts.

This dissertation study presents linguistic analyses of 25 second- and 27 fourth-grade ELLs' argumentative writing produced as part of a Systemic Functional Linguistics (SFL)-informed instructional unit aimed at providing explicit support for children's knowledge of the structural, logical, and linguistic features of argument. An SFL-based approach was used to analyze student arguments at the clause and stage (e.g. Claim, Evidence, Reason, Counterargument) levels. Students' arguments and pre- and post-instruction argumentative writing were also evaluated by independent raters using a primary trait assessment. Independent scoring of pre- and post-instruction tasks showed that the arguments of a majority of students in each grade improved after participation in the argument unit.

The findings of this research indicate that second and fourth graders were able to use developmentally appropriate language features in accordance with expectations for

the genre. Performance across and within stages was generally not predicted by academic or language proficiency level. These results demonstrate that even students whose age is typically regarded as an impediment or who may otherwise have difficulties were capable of constructing effective arguments when supported to do so.

While young children are often supported to attend to such features in writing as voice, organization, and mechanics, this study suggests that they can also benefit from stage-specific instruction and evaluation. The results of stage analysis were thus used to develop grade level-specific descriptions of the range of performance for each stage and corresponding recommendations, including an evaluative framework, for providing targeted pedagogical support for the features of argumentation with which students tended to struggle.

CHAPTER 1

Introduction

The direction in which we are moving is clear: As of this writing, nearly all U.S. states have adopted the Common Core State Standards (CCSS), which promote argumentative writing as a cornerstone of K-12 education, to be taught from the earliest grades on (Common Core State Standards Initiative, 2010). This is a laudable position—and a goal that we are currently unprepared to meet.

Though few would question the importance of writing skills to academic success, children's writing has until recently received far less attention in literacy research than have issues related to reading (Farnan & Dahl, 2003). Research on elementary students' writing has featured a decades-long emphasis on process approaches (e.g., Flower & Hayes, 1981; Scardamalia & Bereiter, 1983; for practitioner-oriented work see Graves, 1983; Humes, 1983). Such work has produced a helpful framework for the start-to-finish development of individual student compositions, and process writing still figures prominently in the way writing is taught to young children in the U.S. (Gilbert & Graham, 2010; Graham, McKeown, Kiuahara, & Harris, 2012). The process writing framework alone, however, does not provide sufficient support for the development of genre-specific knowledge, and while writing research generated over the last several decades has identified a number of practices associated with effective writing instruction, including some work on the teaching of argumentation (e.g., see Graham et al., 2012),

questions regarding how best to support argumentative writing at the elementary level in particular still remain. Given the emphasis on argumentation in the CCSS, expanding our knowledge base in this area has the clear potential to positively impact students' academic achievement.

Argumentative writing, which this study investigates, has long been a highly-valued and high-stakes academic genre for older students, and its importance is reflected in the amount of research that has been done on argumentation at the postsecondary level. Elementary classrooms, on the other hand, have traditionally privileged narrative genres, frequently to the detriment of expository genre development (Duke, 2000; Kamberelis, 1999; Martin, 1989; Pappas, 1993), and this emphasis is similarly reflected in research on elementary students' writing. While the widespread adoption of academic standards emphasizing argumentation across the grades means that argument writing will become increasingly important for students well before they reach the postsecondary level, there have been few studies exploring the abilities and needs of very young students as relates to instruction in argumentation (Newell, Beach, Smith, & VanDerHeide, 2011). The resultant lack of research-based knowledge unfairly leaves practitioners to depend on their "best guesses" in covering the ground between broadly written performance standards and their students' unique needs.

Argumentation involves sets of structural, logical, linguistic, and social features that are distinct from narrative and other types of writing (Freedman & Pringle, 1984; Hillocks, 1999, 2010; Kuhn, 2005), and we currently have an incomplete understanding of how children's argumentative writing abilities can be developed. Children's capacity for argumentation has, I contend, been underestimated in writing research, due in part to

the absence of appropriately supportive pedagogical approaches. Data collected in standardized test settings and traditional instructional contexts paint a dismal picture of student performance. Recent national testing results, for example, show that only 15% of twelfth graders were marked as proficient in writing well-organized essays in which they took and consistently supported positions (Perie, Grigg, & Donahue, 2005), adding another data point to a long trail of troubles with argumentation (see also NAEP, 1994, 1999). Other studies have shown that the quality of written arguments increases with grade level (Knudson, 1992b; McCann, 1989), making such poor performance from students on the cusp of completing their K-12 education even more discouraging. But the picture is incomplete: it is impossible to speak to the limits of student potential without having made a concerted effort to develop it. In order to establish more accurate, evidence-based expectations for student performance, children's argumentative writing needs to be studied in the context of instruction that specifically targets this genre.

While the few instructional interventions that have been implemented in K-12 settings (e.g., Harris, Graham, & Mason, 2006; Reznitskaya, Anderson, & Kuo, 2007; Yeh, 1998) suggest that children benefit from explicit instruction in the structural components and social purposes of argumentation, one aspect that has received little attention is the role language plays in building an argument. This is not surprising; if children's writing has been underrepresented in literacy research, the role of language in literacy development has fared even worse. Despite insistence from prominent literacy researchers and linguists that all teachers need knowledge of how language functions in educational settings (e.g., Wong Fillmore & Snow, 2000), we are largely missing the research and pedagogy that would help teachers understand and explain in meaningful

ways how language actually works in writing. Academic language, oral or written, at any grade level, differs from the conversational registers through which students navigate their social worlds; as students move through the grades and into different subject areas, they encounter increasing abstraction and the complex patterns of language use and organization associated with specialized registers (Schleppegrell, 2004). To the extent that language has been studied, however, instructional research has generally focused on a single component—vocabulary—which, while clearly one part of the knowledge students need, is insufficient to support children’s literacy development when taught in isolation (Anstrom et al., 2010).

Access to knowledge about how language works is essential for all developing writers, and especially so for English language learners (ELLs), whose academic success depends on it. ELLs are a student population that has never been particularly well-served and that is now one of the fastest-growing in the country: Yearly increases in the number of bilingual students have averaged 8% since 1979, a rate that is projected to increase; at the turn of the twenty-first century, 17% of 5-to-24-year-olds spoke a language other than English at home (García & Cuellar, 2006). Several recent reviews of research on instruction for ELLs have concluded that to be effective, teachers of language minority students and ELLs in particular must be able to provide explicit instruction in academic language (August & Shanahan, 2006; Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006; Genesee, Lindholm-Leary, Saunders & Christian, 2006). Specifically, research suggests three main areas that need to be developed to enhance teachers’ support of their ELL students: foundational knowledge about how academic language functions in classroom texts; macro-scaffolding that maps out challenging yet well-supported goals for language

and content learning over time; and micro-scaffolding supporting the moment-to-moment unfolding of instruction in the classroom (Schleppegrell & O'Hallaron, 2011).

While currently available pedagogical approaches may allow teachers to take an in-depth look at individual components of literacy instruction—vocabulary instruction, for example—almost all lack a robust theory of language that would help educators make conceptual connections across the different parts of literacy curricula. A notable exception is Halliday's (1994) Systemic Functional Linguistics (SFL), which offers a framework and a corresponding metalanguage, or way of talking about language, for making the social, structural, and linguistic aspects of texts visible and accessible to teachers and students. Pedagogical approaches in this tradition, frequently built around the exploration of school-based genres, have in recent years become more visible in the U.S. schooling landscape, and teachers working under this model report significant improvement in their ELLs' engagement with and production of targeted text types (Brisk & Zisselsberger, 2010; Schleppegrell & de Oliveira, 2006; Schleppegrell et al., 2013; Gebhard, Harman, & Seger, 2010). SFL approaches to argumentation have not yet been extensively studied in U.S. elementary schools, but related research suggests the potential for explicit, genre-based writing instruction to meet teachers' need for concrete pedagogical tools.

The present study presents a detailed analysis of the argumentative writing of second- and fourth-grade English learners whose learning of that genre has been highly scaffolded through an SFL-based approach called Functional Grammar. Description of classroom instruction, discourse, and artifacts will make visible the interactional elements of the richly supportive pedagogical approach through which students' learning was

developed. Taken together, examination of children's arguments and the instructional environment in which they were produced will enhance understanding of what English learners at different grade and language ability levels are capable of producing when their learning takes place in supportive contexts; how children learn to produce written argumentation; and how we can better support them to do so.

This research, then, responds to calls to make the linguistic and text structures of academic language visible to ELLs (Gibbons, 2002; Snow & Kim, 2010); to study the use of metalanguage as a means of doing so (Purcell-Gates, Duke, & Martineau, 2007); to closely examine aspects of genre knowledge development in schools (Gebhard & Harman, 2011); and to investigate specific instructional methods for the teaching of argumentative writing and how children make sense of what they learn (Newell, Beach, Smith, & VanDerHeide, 2011).

Present Study

If we hope to develop an accurate understanding of children's argumentative writing capabilities, observations based on young students' unsupported writing are insufficient. To speak to the potential for these students to develop effective arguments, we instead need to know what kind of writing students produce in appropriately supportive instructional contexts. To that end, this dissertation study presents analysis of second- and fourth-grade English learners' argumentative writing produced as part of an instructional unit aimed at providing explicit support for children's knowledge of the structural, logical, and linguistic features of argument. The unit featured opportunities for students to explore the way sample texts unfolded and how each step contributed to their authors' achievement of their overall goals; examine the role key language features

played in conveying meaning; enact argumentation through discussion and debate with their peers; and build their own arguments piece by piece, reflecting on their thinking and writing in whole-group and small-group conversations as they went.

Subsequent chapters present analysis of second and fourth graders' arguments at the level of the clause and the *stage*, or structural element. Clause-level analysis demonstrated that the writing of children in both grades contained features that were both developmentally appropriate and consistent with genre-specific goals. Stage-level analysis revealed that second and fourth graders were, in fact, able to effectively engage with and express the types of reasoning characteristic of the component elements of argumentation. This finding challenges the prevailing beliefs regarding the argumentative abilities of second graders in particular, who are often either assumed to be less capable of producing complex arguments than their older peers—or, as is more often the case, excluded from study altogether. In addition, comparison of performance between academic and language proficiency level showed that, counter to what one might expect, neither was predictive of the effectiveness of students' arguments. In the context of the instructional unit described here, the quality of arguments was instead largely a matter of individuals' strengths and weaknesses.

Examination of student texts showed that these strengths and weaknesses varied not only between students but also between elements of argumentation. The study's fine-grained linguistic analysis yielded specific observations regarding the precise areas in which students tended to struggle and the ways in which these struggles manifested themselves in writing. Such observations, and the resultant pedagogical recommendations, are a key contribution of this work to research on children's

argumentation. Research in this area often assesses children's writing using holistic measures that can speak broadly to the quality of written arguments but are incapable of providing detailed explanations of *why* students may struggle with a particular element. Attention to the linguistic aspects of argumentation, in particular, is often entirely absent in evaluative instruments. Characterization of the role that language features play in shaping the effectiveness of young students' writing will help researchers and practitioners provide targeted instructional support for features that are likely to contribute to difficulties; it also raises the question of whether difficulties that have been attributed by some to cognitive immaturity (e.g., Freedman & Pringle, 1988) are in fact attributable to language issues that can be effectively addressed through supportive pedagogy.

This study also compares the results of linguistic analysis with the scores assigned by teachers recruited to evaluate writing using a traditional primary trait rubric. Independent raters scored students' arguments and pre- and post-instruction argumentative writing tasks. Scoring of the latter tasks showed that a majority of students in both grades improved on some or all of the traits measured by the rubric. Independent scoring of the arguments produced as a result of students' participation in the argument unit provided further support for the findings of linguistic analysis, while also revealing important differences in the values that guide each approach to assessment. Finally, a major contribution of this research is the elaboration of a framework for the evaluation of argumentative writing. The framework provides concrete, stage-specific guidelines to assist teachers in evaluating writing and generating fine-grained, individualized feedback,

and in doing so brings together a number of facets of fact-based argumentation that existing rubrics and analytic frameworks do not integrate.

In exploring the issues described above, this work addresses the following research questions:

1. What were the features of students' argumentative writing at the second and fourth grade level when supported by a Functional Grammar pedagogy? How do the features of students' arguments differ by academic achievement or language proficiency levels?
2. What were the Functional Grammar-informed argument curriculum and other features of the instructional context in which English learners' written argumentation in second and fourth grade was observed? How was this reflected in the features of student writing? Does scoring of pre- and post-instruction writing using a primary traits rubric show that students' arguments improved after participating in instruction?
3. What relationship, if any, exists between students' performance as measured by linguistic analysis of the structural components of their arguments and the scores they receive on a traditional primary trait assessment? How do these approaches to assessment differ in terms of the potential of the feedback they generate to inform development of argumentation skills?

These questions are addressed in six chapters. Chapter 1 introduced the study and described the problems it aims to address. Chapter 2 presents a review of the literature relevant to the present study. It begins with a discussion of studies of argumentative writing in K-12 contexts, including studies of children's argumentative writing produced in traditional instructional contexts and as a result of instructional interventions. Then it

provides an overview of Systemic Functional Linguistics, the social theory of language underpinning this study, and presents examples of how this theory can be applied to the study of young students' writing.

Chapter 3 presents the study's methods. It describes the larger project of which this research was a part; context and sampling considerations; and the rationale for selection of a traditional primary trait assessment. It also outlines the principles guiding the design of the pedagogical approach used in the study.

Chapters 4 and 5 contain analysis of data collected in the second and fourth grade, respectively. They follow the same structure, each providing detailed descriptions of how instruction unfolded over the course of the argument unit; reporting on scoring of pre- and post-instruction writing; presenting fine-grained linguistic analysis of student writing at the level of whole class, academic and language proficiency subgroups, and focal students; and discussing results of independent scoring using a primary trait rubric.

Finally, Chapter 6 provides a synthesis and comparison of second- and fourth-grade findings and discusses limitations and implications of the study.

Chapter 2

Review of Literature

This work addresses the need to develop empirical descriptions of children's argumentative writing that will help inform research-based argument writing pedagogy. Accordingly, this chapter surveys the literature pertaining to children's written argumentation and the role explicit instruction can play in developing the logical and linguistic aspects of their argumentative abilities. It also provides an introduction to Systemic Functional Linguistics, the theory of language and learning that guides the present study, and presents examples of how this theory can be applied to the study of children's argumentative writing. In doing so, it argues that additional research into the argumentative capabilities of very young students as well as students at varying levels of academic achievement and English language proficiency is needed, and that an accounting of the linguistic aspects of learning about and writing arguments is an understudied and sorely needed component of such research.

Argumentative Writing

Given the weight historically placed on argumentation at the post-secondary level, it is unsurprising that argumentative writing is relatively well-represented in first- and second-language research. This work, however, has limited application for learners in the primary grades. While there is also a fairly established body of cognitively-oriented

research on supporting elementary students' argumentative reasoning skills through group discussion in traditional classrooms (e.g., Anderson et al., 2001; Jadallah et al., 2011; Michaels, O'Connor, & Resnick, 2008) and particularly in science classrooms (for a review, see Cavagnetto, 2010), it is important to note that writing instruction is generally not a central feature of these designs, which leaves students who have not already learned how to carry out the appropriate discursive moves at a disadvantage. Studies that do examine children's written arguments show that performance tends to be poor in traditional instructional contexts. At the same time, they make clear that with adequate support, young students are able to engage productively in this kind of reasoning. The following section addresses definitional issues, then reports on research on children's argumentation in traditional instructional contexts and, separately, as the product of instructional interventions.

Argumentation, persuasion, or opinion piece? In research and practice, the type of writing under study here goes by many names. At times the distinction between the labels *argumentation*, *persuasion*, and *opinion* writing are meaningful, representing differences in what a piece of writing hopes to accomplish and how. At other times these labels are used interchangeably. The *opinion piece* label appears most often in elementary contexts, including in the Common Core State Standards descriptions for writing in Kindergarten through Grade 5 (Common Core State Standards Initiative, 2010a). While all of the basic elements of an argument are expected to be in place by the end of fifth grade, it is not until sixth grade that the CCSS begins to use the term *argument*. The Standards' authors indicate that the use of the label *opinion* for the elementary grades in

the CCSS in particular is meant to suggest a pared-down, developmental form of argument—it is, in other words, more a matter of scope than substance.

On the other hand, researchers, practitioners, and curriculum designers tend to differentiate between persuasion and argumentation on the basis of qualitative differences. Textbook authors Lunsford and Ruszkiewicz characterize argumentation as being about discovery, while persuasion is about changing another’s point of view or inciting them to act. They summarize the contrast by saying that “writers or speakers argue to find some truth; they persuade when they think they already know it” (2004, p. 7). Fulkerson (1996) further clarifies that where persuasion is a unidirectional attempt at convincing the reader of a single correct answer, an argument is a dialectical exchange, the purpose of which is to reach wise decisions; these decisions, he adds, are always subject to revision in light of better arguments or evidence. Others (e.g., Derewianka, 1991) have differentiated between these purposes using the terms “persuading *that*” and “persuading *to*”.

The authors of the CCSS concur with these broad distinctions, as they make clear in an explanatory appendix. To help stakeholders understand the Standards’ emphasis on *argument* rather than *persuasion*, they write:

When writing to persuade, writers employ a variety of persuasive strategies. One common strategy is an appeal to the credibility, character, or authority of the writer....Another is an appeal to the audience’s self-interest, sense of identity, or emotions....A logical argument, on the other hand, convinces the audience because of the perceived merit and reasonableness of the claims and proofs offered rather than either the emotions the writing evokes in the audience or the

character or credentials of the writer. The Standards place special emphasis on writing logical arguments as a particularly important form of college- and career-ready writing. (2010b, p. 24)

Given the inconsistency in labeling from one context to the next—and the implications of these differences in terms of purposes and component skills associated with a given writing task—it is necessary in surveying the literature to pay close attention to how authors define the constructs with which they are working. To that end: The text type presented in the instructional approach at the heart of this study is not consistent with the aforementioned descriptions of *persuasion*. Furthermore, in the following chapters I will present evidence to counter the Standards writers’ claim that “young children are not able to produce fully developed logical arguments” (2010b, p. 23), making the *opinion* label a poor fit as well. The appropriate label for the kind of writing discussed here, then, is *argument*¹, and that term will be used in describing this study. Perhaps not surprisingly, this terminology brings with it another set of definitional issues, to be discussed in the next section.

Defining argumentation. The model of argumentation guiding the present study emphasizes the social purposes of argumentation and the structural and linguistic means by which those purposes are achieved. It is informed by Systemic Functional Linguistic theory, to be discussed in greater detail below, but also shares a number of features with models used in other studies of written argumentation.

¹ The text type in question is referred to in the literature as *argument*, *argumentation*, *argumentative writing*, and even *argumentational writing* (e.g., Andres, Torgerson, Low, & McGuinn, 2009). These terms vary according to author preference and are mostly interchangeable; the first three terms listed here will be used as such in this document.

Most research in argumentative writing draws on the three-part model elaborated by Toulmin (1958) to some degree. While the model recognizes that some aspects of what is valued in argumentation vary by field, its basic structural elements do not: they are considered to be field-independent. This basic model consists of a *claim*, *evidence* (alternately labeled *data* or *grounds*), and *warrants*. The claim element is defined by Toulmin as “conclusions whose merits we are seeking to establish” (1958, p. 97). Evidence include facts or examples supporting the claim. Warrants are general, law-like statements that act as bridges between evidence and claim. In Toulmin’s words:

“Our task [in generating warrants] is no longer to strengthen the ground on which our argument is constituted, but is rather to show that, taking these data as a starting point, the step to the original claim or conclusion is an appropriate and legitimate one.” (p. 98)

Toulmin observes that warrants are frequently implied rather than explicitly stated. Indeed, their often-implicit nature can make them difficult for students to identify, presenting a challenge for teachers of argumentative writing well into the secondary and postsecondary years (Rex, Thomas, & Engel, 2010; Warren, 2010). In the oral arguments of younger children, warrants are typically missing altogether (Reznitskaya, Anderson, & Kuo, 2007). Nonetheless, warrants are crucial to the success of any argument, and for this reason an important aspect of the work presented here involves helping children articulate the warrants underlying their written arguments.

Toulmin further elaborated his model through discussion of additional elements that may be included in argument, including *backing*, *qualifications*, and *rebuttals*. Backing elaborates on and provides further support for warrants. Qualifications are

expressions that limit the strength of arguments, and rebuttals anticipate potential counterarguments or outline conditions under which the claim would not hold true. Secondary constructs such as backing or qualifications do not tend to feature prominently in research applications, and many researchers expand their descriptions of argument structure to specifically include a more direct presentation of counterarguments than Toulmin's model suggests. These observations are also true for the present study.

While Toulmin's work serves as a common starting point, argumentation is a complex construct for which there is no single agreed-upon definition or description in the relevant literature. The evaluation of arguments for investigative purposes requires that researchers articulate what they consider the basic elements of argumentation to be, and also that they establish criteria for judging the quality of those elements. Standards for quality—in other words, the characteristics of an ideal argument—are defined by different researchers in different ways. These definitions vary in their emphasis on the importance of structural elements, relationships between ideas, and the quality of ideas themselves to the overall success of the argument. For example, Freedman and Pringle (1988) used minimal structural standards compared to Toulmin, instead focusing heavily on the relationships between structural components. They stipulated that an argument consists of a single thesis and multiple propositions, all of which must be located “within a hierarchic structure so that each proposition is logically linked not only to the preceding and succeeding propositions but also to the central thesis and indeed to every proposition within the whole text” (p. 234). In this view, based on traditional views of formal logic, the organization of premises is paramount, while the ideas themselves are given second billing. Voss and Means (1991; see also Means & Voss, 1996), meanwhile, also used a

fairly simplified model of argument that included at minimum a claim with one supporting reason, and ideally a counterargument with a supporting reason. In addition to these simple structural standards, their study of informal reasoning (i.e., everyday arguments that are not necessarily bound by the rules of formal logic) used three criteria to determine the soundness of an argument, focused on analyzing the quality of the content: the general acceptability of the reason(s) given, the relevance and strength of support the reason provides for the claim, and the extent to which counterarguments are taken into account.

This study's view of what constitutes sound argumentation integrates the three perspectives outlined above. The social act of carefully weighing information to reach a conclusion and make a recommendation is seen as unfolding through a sequence of field-independent structural components; the relationships between these components are seen as instrumental in effectively conveying the writer's ideas; and the strength and validity of the ideas themselves are, of course, essential in constructing meaning. Some combination of these perspectives is common, but in the same way that it is important to look closely at what researchers intend to signal through their choice of label for the text type under study, the use of the label "argument" should be carefully examined, as the way it is defined (or not) provides meaningful information about the assumptions and values guiding the work that informs what we think we know about children's argumentative writing. A survey of the relevant literature suggests that some key assumptions may need to be revisited.

Research on argumentative writing. While there is a substantial research base concerning argumentative writing at the postsecondary level, the available studies of

children's written argumentation are quite limited in number, especially as concerns elementary-age students. The following section reports on argumentative writing in K-12, primarily North American, contexts. Studies from general education contexts in which students reason based on their own opinions and experiences, literature, or informational texts relevant to a particular topic are included. Not included is research specifically directed at using argument to foster scientific practice and/or scientific literacy (for a detailed discussion and review of argumentation in science contexts, see Cavagnetto, 2010); while the present study does draw on informational text in support of an argument, its core focus is the development of argumentative writing itself rather than the discipline-specific goal of using argumentation as a means of exploring scientific phenomena. A group of studies characterizing the arguments produced by students in traditional instructional contexts (i.e., without specific support for argumentative writing) is presented first, followed by another set of studies examining arguments produced in the context of instructional interventions. Finally, issues in evaluation that emerge from the surveyed research are briefly discussed.

Traditional instructional contexts. Much research on argumentative/persuasive writing has focused on what children produce without being exposed to purposeful instruction in argumentation beyond what local curricula would normally include. For example, McCann (1989) performed a study of 95 students in grades six, nine, and twelve. McCann developed a guide to evaluate the quality of six argumentative traits: Claims, Data, Warrants, Proposition, Recognition of Opposition, and Response to Opposition. The latter two elements, as their names suggest, were modifications of the Toulminian model that brought writers' treatment of counterarguments into sharper

focus. Propositions, which were also not included in Toulmin's model, were essentially suggestions regarding actions that could be taken to address the issue presented in the writing prompt used for the study. McCann found that the younger students in his study performed less well than their older counterparts. Sixth graders, for example, showed almost no use of warrants, while older students used them more often. Sixth graders also scored significantly lower than older students when it came to recognizing and responding to possible opposing arguments. Grade level was not a surefire predictor of success, however, as students at all levels were found to have difficulty with some features of the genre.

A later study by Knudson (1992a) produced similar findings. Knudson collected argumentative writing samples from a total of 202 children in grades four, six, ten, and twelve. No data pertaining to instruction were collected. Knudson analyzed the samples using the rating system produced in McCann (1989) along with a holistic scoring guide meant to provide a more general sense of students' writing competence. She described the general characteristics of writing at each of the four grades with respect to the relative frequency of use of Toulmin's traits, and concluded that, as might be expected, fourth-grade writers produced poorer arguments than did older students, and performance on both argument-specific and holistic measures increased with grade level. Claim and propositions were found to be the elements that students across grade levels were most effective at writing. Notably, although both McCann and Knudson used the *argument* and *persuasion* terms interchangeably, the writing prompts used in Knudson's study were written in such a manner that they would be likely to produce writing consistent with the more specific definition of persuasion described in the previous section of this chapter.

Students in her study were explicitly asked to convince their target audience (the school principal) to take a particular course of action, and this difference could well account for the increased presence and/or effectiveness of propositions in students' writing.

Additional evidence for the long-term development of familiarity with argument structure in the absence of explicit instruction comes from a pair of reading studies carried out by Chambliss (1995) and Chambliss and Murphy (2002). Chambliss's (1995) study found that text structure consistently influenced twelfth graders' comprehension of written arguments, and that by this point in their schooling students were adept at using the claim-evidence-warrant relationship to identify whether a text was an argument or another type of informational text, but they were least sensitive to warrants. (Where participants were able to identify warrants, they had no vocabulary to describe them, referring to them in some cases simply as the author's opinion.) Chambliss and Murphy's (2002) follow-up study of fourth- and fifth-grade children's representation of the global discourse structure of argument found that the text cues that had been salient for older readers were less so for younger children, but that the elementary school students were nonetheless able to identify a hierarchical global discourse structure (i.e., a superordinate claim and supportive data points). Though differences between grades were not statistically powerful, their findings also suggested that fifth graders were more able to represent the global structure of text they had read, and to infer implicitly stated claims, than fourth graders. While these comprehension studies are not predictive of students' productive abilities, taken together they give a sense of baseline familiarity with argument structure at different ages.

In an attempt to explain why students consistently struggle to write arguments, Freedman and Pringle (1988) examined the narrative and argumentative writing of 500 students in Canadian grades seven and eight. As described above, their evaluative criteria included minimal requirements for specific structural elements but emphasized the logical relationships between elements. They found that, while over 98% of students wrote stories evidencing conventional narrative structure, only 12.5% of the same students were able to realize the minimal criteria specified for argumentation. By way of explanation, Freedman and Pringle noted that the children had minimal exposure to written arguments either at school or at home, especially as compared to their experiences with narrative texts. The larger part of their discussion, however, centered on the idea that poor performance on the argumentative task could be accounted for by insufficient cognitive maturity on the part of the writers. This theory, while shared by some other researchers, (e.g., Bereiter, 1980; Felton & Kuhn, 1997) could not be substantiated using the data the authors collected for their study.

Crowhurst (1990a) reported on the argumentative writing of students in Canadian grades 5, 6, and 7 and offered another potential explanation for students' difficulties. Based on her analysis of some 1200 persuasive compositions, she characterized young writers' arguments as typically relatively brief, with simple claims, "baldly stated, unelaborated reasons that often sound like a list" (p. 351), and lacking concluding statements. Writing was usually not logically organized; rather, students in the sample tended to produce associative writing, with each sentence relating to the one before it. At times the writing produced in response to persuasive/argumentative prompts was not typical of the appropriate genre. For example, some students began writing an argument

but then shifted into writing that was narrative in function, while others responded with dialogues or a separate kind of text Crowhurst refers to as an “informative composition.”

Of interest for the present study, Crowhurst made special note of the language features of children’s arguments, suggesting that differences in performance from one grade level to the next could also be attributable to linguistic factors. Specifically, she reported that writers in her sample showed a lack of diversity and precision in vocabulary, repeatedly used a limited number of connectives, and wrote short sentences with structures more typical of oral language. Based on her previous studies of student writing (Crowhurst, 1980, 1987, 1990b), Crowhurst observed that argumentation “characteristically has longer clauses and T-units than narration, is typified by more complex constructions such as nominalizations, and depends on logical connectives to signal relationships between sentences” and that ability in all of these areas increases with age (p. 355). Younger students’ difficulties could thus be accounted for in part by their still-developing ability to express themselves through language rather than a cognitive incapacity for argumentation. Developmental issues notwithstanding, Crowhurst noted that even the younger students in her sample displayed rudimentary understanding of some features of argumentation, and recommended that practitioners engage younger children in reading and writing arguments to develop these abilities.

Crammond (1998) collected persuasive writing from sixth, eighth, and tenth graders and analyzed in detail 12 pieces at each grade level and compared them to the writing of seven experts (professional writers who routinely wrote argumentative texts for publication). The writing of each group was characterized in terms of the argument structures used and their complexity. Crammond’s model of argument was the most

comprehensive of the studies presented here, including all of the elements (“substructures”) of Toulmin’s model and even expanding on some of their descriptions. To aid in identification of this complex network of argumentative substructures, she mapped the relationships between each substructure and the general semantic structure she identified as being associated with it. These semantic structures and particularly their characteristic conjunctions were used in coding the writing. Crammond concluded that “the ability to produce a basic written argument is acquired at a relatively early age—at least by Grade 6” (p. 249), though, as might be expected, the use of embedded and elaborated argument structures increased with age. In particular, the youngest students in the study included no warrants and very few “countered rebuttals”, which Crammond interpreted as an indication that younger students were less audience-focused in their writing. She suggested that, in order to promote the effective use of these substructures, teachers should support students in rhetorical skills related to audience analysis.

Finally, the work of Anderson and colleagues (1997) provides evidence of young children’s argumentative capabilities by focusing on oral, rather than written, arguments. Seeking to describe the properties of children’s naturally occurring arguments, the authors analyzed conversations from four fourth-grade classrooms. The conversations took place as groups of five to ten children shared with one another their arguments related to dilemmas posed by stories read in class. As they discussed their positions, students frequently made seemingly ambiguous references, and claims and/or warrants were often inexplicit or missing. Analysis of classroom transcripts, however, demonstrated that these ambiguities were nearly always easy to understand or reconstruct in the context of the group conversation. The authors argued that, rather than judging

young students' arguments as logically lacking, contributions by cooperative participants should be taken into consideration when evaluating their oral arguments—in other words, that an individual's choice to not restate certain elements of argument is actually entirely sensible and appropriate in such a collaborative setting. These findings have important implications for written tasks, which, in academic settings, typically require each student to independently justify their arguments to a distant audience. If, in fact, classroom arguments do not ask children to make their reasoning clear on an individual basis, distinct from group conversation, this is all the more reason to expect that they will need explicit support to do so in a medium—writing—that requires it.

Taken together, these studies corroborate national data on argumentative writing performance (e.g., NAEP, 1994, 1999), showing that in traditional instructional contexts students' performance on argumentative tasks improves over time but remains below standards for many students even at the highest grade levels. Theories suggesting cognitive or linguistic causes have been advanced, but the power of any such explanation is limited in the absence of focused instruction in argumentation, without which little can be said about what students might truly be capable of.

Argument-specific instruction. Another group of studies examines students' writing as a result of argument-specific instructional interventions. These studies paint a more optimistic picture, and provide insight into the instructional features that are likely to be effective in supporting argumentative performance.

Interventions in middle school through university. Research on instructional approaches to support older students' argumentative writing, for example, demonstrates the potential benefits of explicitly teaching argument structure and purpose. Varghese and

Abraham (1998) provided Singaporean university students with eight weeks of explicit instruction in these aspects of argumentation. The researchers' model of argument included Toulmin's claims, grounds, and warrants, and they defined argument as "discourse enacted to produce shifts in beliefs, effected through logical or emotional appeals" (p. 289). Explicit instruction in this case involved familiarizing students with the basic elements of an argument; providing opportunities to use those concepts to evaluate the strength of arguments which varied in quality (amounting to three or four such oral critiques per week); and, finally, having students write their own arguments. Students wrote several shorter and longer pieces (500 and 1000 words, respectively), using Toulmin's model to evaluate their own work. The same argumentative task, chosen for authenticity and familiarity, was used for both pre- and post-tests. Responses were scored using three-point scales for each structural element, and statistically significant improvement was observed in the quality of students' claims, evidence, and warrants.

At the middle school level, Hidi, Berndorff, and Ainley (2002) tested the effectiveness of an eight-week intervention designed to provide motivational and instructional support for sixth graders' argumentative writing, including discussion in four elements of argumentation: claim, evidence, rebuttal, and conclusion. Here too, holistic scores for all students increased between the pre- and post-test. Other key findings come from Yeh (1998), who investigated the effects of explicit instruction and pre-writing heuristics on middle school students' argumentative writing. He designed a six-week unit for argumentation during which all students in four seventh-grade language arts classes were immersed in reading, debate, peer response, and writing activities. An experimental group formed by a subset of participating classes also received instruction

in the use of two pre-writing heuristics based on Toulmin's model. Analysis of the experimental group's writing showed larger pre- to post-test gains in development and voice, and interviews with focal students indicated that those in the experimental group had significantly higher knowledge of argument criteria and strategies. Furthermore, Yeh examined students' application of the heuristics and found that a "substantial percentage...applied the heuristic fully and flexibly" (p. 68)—in other words, that the explicit instruction did not lead to rote or rigid reproduction of what was taught, a common concern for those who are suspicious of explicit approaches. While the interventions described in these three studies vary somewhat in their design, an important commonality is that none of them rely solely on discussion of structural elements; all are dynamic, providing students opportunities to read, discuss, and write arguments on different topics over the course of weeks or months. Yeh's (1998) findings are particularly important because they demonstrate the added benefit of explicit instruction beyond the gains one might expect for students in the otherwise identical immersion environment. Interestingly, and of relevance to the present study, effects were particularly strong for arguments written by cultural minority students.

A description of apparently less effective approaches comes from Knudson (1992b), who investigated the effectiveness of four "instructional strategies" (p. 169) on tenth and twelfth grade students' argumentative writing. The strategies included presentation of model pieces of writing; presentation of scales, questions, and criteria to guide writing and revision; models combined with scales/questions/criteria; and free writing. (The scales, questions, and criteria approach focused students on matters of audience and opposing arguments.) Scoring of essays produced after 14 days of

instruction revealed no significant difference between the strategies in terms of effectiveness at either grade level. Notably, though Knudson drew on Toulmin's model to score the essays, none of the treatments involved exposing students to information about the structural elements of arguments. While the instructional approaches investigated in the study may be beneficial in combination with other strategies, the findings from this study make it clear that exposure to models and focusing on interpersonal aspects of argumentation were insufficient on their own.

Elementary school interventions. Studies that examine explicit instruction at the elementary level provide cause for optimism. A number of high-quality studies on children's argumentation have resulted from investigation of the Collaborative Reasoning (CR) framework, which emphasizes the dialogic aspects of argumentation. CR provides students with practice in putting forth their arguments, supporting them with evidence, and making or responding to counterarguments. A particularly interesting example is Reznitskaya, Anderson, and Kuo's (2007) quasi-experimental study, designed to measure the processes underlying the development of argumentative knowledge. A previous study (Reznitskaya et al., 2001) had found group discussions using the CR model significantly improved students' arguments. The 2007 study sought to determine whether performance could be further improved through explicit instruction on the characteristics of argument. 128 fourth and fifth grade students from two schools were assigned to one of three conditions: 1) regular reading instruction, 2) four CR group discussions, or 3) four CR group discussions plus two scripted lessons on argumentation, given after the first two group discussions had taken place. The lessons described the definition, purpose, and uses of argument and laid out the five parts of the researchers' basic

argument schema: position, reasons, supporting facts, objections, and responses to objections. (Warrants were omitted entirely based on the researchers' desire to focus on "the most crucial discourse elements that are already present in the arguments of young children or that can be introduced through developmentally appropriate instruction" (p. 454); the latter part of this view is contested by the present study.) At the end of the two remaining group discussions, the students classified parts of those conversations according to the elements of the argument schema in order to reinforce their learning. Though writing was one of the study's outcome measures, writing instruction itself was not a central feature of the design, and no mention is made of opportunities to practice writing using the schema.

The post-intervention task of most interest here was a reflective essay, intended to assess how children were able to apply what they had discussed in a group setting to their own writing. Of critical importance is the authors' stated interest for evaluating writing, which was in students' ability to express an internal dialogue (rather than the persuasive power of the writing or ability to use the conventions of written argument). Accordingly, essays were evaluated on a) the number of unique, argument-relevant reasons given and b) the counterarguments addressed. Contrary to the researchers' expectations, students in the explicit teaching condition scored lower on their essays, as they gave fewer unique argument-relevant propositions in their writing than did students in the discussion-only treatment. This should not have been entirely surprising: As the authors note, the writing task became more structured and cognitively demanding for these students as a result of their additional knowledge. The dimensions valued by the study design may have favored the discussion-only group, whose writing was not constrained by attempts to apply the

argument schema, which modeled inclusion of just two reasons. In fact, while they included more reasons, writers from the discussion-only group “were much more verbose, often suffering from a lack of planning or critical monitoring” (p. 465). This suggests that the explicit instruction condition provided benefits that the study was simply not designed to capture, as the authors’ focus on internal dialogue would not reveal whether explicit instruction resulted in, for example, arguments that were shorter but more logically sound. It also bears mentioning that these results were based on participation in just two lessons; given that argument is well-established as a particularly demanding genre for students of all ages, it is not surprising that gains would be limited under these circumstances.

Another instructional model that has been tested at the elementary level is Self-Regulated Strategy Development (SRSD; for detailed descriptions, see Harris & Graham, 1996, 1999). SRSD, which has been used for instruction in a number of written text types, targets students’ strategic behavior, domain-specific knowledge, and motivation. Students are provided with strategies for planning, carrying out, and reflecting on writing tasks, including both general and genre-specific writing strategies. Following studies demonstrating the effectiveness of SRSD in upper-elementary and middle school (Graham & Harris, 2003) and third grade (Graham, Harris, & Mason, 2005), Harris, Graham, and Mason (2006) tested the program with a group of 63 second graders from an urban school district over a period of several months. The model for persuasive writing used in second grade consisted of a general statement of belief followed by three or more reasons and a concluding statement. (This represented a simplification of a previous model that asked students to explain or say more about their reasons; elimination of this

component was based on the observation that some third graders had struggled with this step in a previous study.) SRSD instruction is scaffolded such that concepts are introduced in a whole-group setting, modeled by the teacher, and applied in a collaborative writing activity before students are asked to apply the strategies in independent writing tasks. The second-grade study also tested a peer support component in which partners were encouraged to help one another apply strategies outside of official instructional time.

Students in the study produced pre- and posttest writing samples. The persuasive writing tasks consisted of questions about their school or home lives, e.g., “Should children be allowed to choose their own pets?” Writing was evaluated using an 8-point genre-specific rubric and was scored for inclusion of basic elements of persuasive writing. Relative to a comparison (Writer’s Workshop) classroom, second graders in both SRSD conditions (with and without peer support) wrote persuasive posttests that were longer, qualitatively better, and included more basic elements of persuasion; additionally, students in the peer support condition included more basic elements than those in the SRSD-only condition. These results provide further evidence that even very young students can benefit from explicit instructional support for persuasive writing, and that purposive peer interaction may be a useful element in such instruction. Importantly, children in this study were classified as struggling writers. Though this study did not compare the performance of struggling students with average or above-average writers, other studies that have done so (e.g., Englert, Raphael, Anderson, Anthony, & Stevens, 1991; Ferretti, MacArthur, & Dowdy, 2000) have found that students with and without

learning difficulties benefit from highly-scaffolded instruction in argumentation, with neither group performing better or worse than the other after instruction.

Unanswered questions. The studies outlined above leave little doubt that young students are able to engage productively in argumentative reasoning when they are supported to do so. There is not yet enough research to determine which, if any, approach is strongest—a study comparing CR and SRSD in a middle school context, for example, found that both were effective in improving students’ written arguments (Coker & Erwin, 2011). Existing studies do, however, offer some insight into the features of instructional designs that are likely to be successful in improving students’ performance. Andrews, Torgerson, Low, and McGuinn (2009; see also Andrews, Torgerson, Low, McGuinn, & Robinson, 2006) carried out a systematic review of the evidence of successful practice as regards teaching argumentative writing to 7 to 14 year olds. Although they only reviewed quasi-experimental or true experimental studies, their findings are consistent with the studies surveyed here. They concluded that successful strategies include “heuristics; planning; oral argument, counterargument, and rebuttal to inform written argument; explicit goals (including audiences) for writing; teacher modeling of argumentational [*sic*] writing; and ‘procedural facilitation’” (p. 291) as well as “peer collaboration, thus modeling a dialogue that (it is hoped) will become internal and constitute ‘thought’; [and] explicit and very clear explanations for students of the processes to be learned” (p. 301). Effective interventions, then, would likely include most or all of these features.

One question that remains outstanding concerns how the specific instantiations of these features might or should change depending on the student population being targeted. For instance, while a handful of studies address instruction for struggling writers

and/or students in urban, low-income schools, the literature is silent when it comes to the argumentative skills of young English learners. Moreover, it is clear that we have very little empirical evidence of the argumentative abilities of very young children in general. Of the studies reviewed above, only one (Harris, Graham, & Mason, 2006) attempted to engage second graders in an instructional intervention aimed at improving their written argumentation; none of the studies analyzing children's writing in the absence of such an intervention included the work of students at this grade level. Indeed, while it is clear that argumentative abilities continue to develop well into the teenage years, there is some disagreement with respect to younger children's abilities to grapple with the demands of producing and comprehending written arguments, especially before the age of about 10. This lower boundary of argumentative ability seems to be widely taken for granted, yet largely unsubstantiated by recent educational research. The present study will address these missing elements and contribute to a research-based understanding of argumentation in second- and fourth-grade English learners at different language and academic proficiency levels.

In particular, this study presents a careful examination of the features of children's writing in order to characterize performance at these two grade levels and identify areas in need of further support. While some research on students' unsupported argument writing (e.g., Crammond, 1998) involves analysis that looks closely at the language typically associated with the types of reasoning inherent in different elements of argument, studies reporting on writing as a product of instructional intervention have thus far not done so. Instead, they have largely relied on numeric scales and holistic scores as measures of students' performance on each element or for arguments as a whole. These

measures are useful in terms of gauging the effectiveness of interventions, but may fall short of providing the level of detail needed to design writing instruction that is targeted to the specific needs of a given student population. Tellingly, descriptions of writing instruction—for those studies that specifically included it—are remarkably similar from one grade level to the next. Some of this is certainly attributable to space considerations, but I contend that it is also attributable to the lack of research that would guide the setting of grade level-appropriate expectations, a problem that this study attempts to address.

Finally, this study also contributes a method of evaluation for researchers and practitioners based on a comprehensive view of argumentation. Research that focuses on only a subset of structural elements—omitting, for example, warrants or counterarguments—has produced analytic tools that are insufficient for those who wish to reintroduce the omitted elements into their instructional or evaluative models. As noted above, rubrics or scales that may be adequate for speaking to student performance more broadly may prove too general to be helpful in generating more specific descriptions. This is especially true for a number of rubrics that focus on logic and logical relationships but fail to address issues specific to written language. On the other hand, tools allowing researchers to produce rich descriptions of student writing exhibit a level of complexity that represents a barrier to entry for practitioners seeking to evaluate their own students' writing or even for other researchers who may not have extensive training in the relevant theoretical or methodological approaches. The evaluative framework produced as a result of analysis in the present study seeks the middle ground, accounting for issues of both language and logic in a moderately elaborated model of argument in a way that is thorough enough to generate useful feedback, yet supportive enough to be useful to

teachers and researchers with basic understanding of the theory that informs it. That theory, Systemic Functional Linguistics, is discussed below.

Systemic Functional Linguistics and Genre Theory

In this study, a Systemic Functional Linguistic (SFL) approach is employed to address the need for research on the role played by explicit writing instruction in children's argumentative writing. The SFL tradition of genre-based pedagogy is particularly important to this work. This section presents an overview of SFL theory and constructs; briefly compares SFL genre theory with other views of genre; and presents two studies that use SFL to engage elementary-level students in written argumentation/persuasion, discussing the contributions they make and how my own work expands upon them.

Systemic Functional Linguistics. Halliday's (1994) SFL is a social theory of language and learning. This means that SFL treats language not as a set of decontextualized rules, but as a meaning-making resource; under this framework, the term *grammar* represents the range of linguistic choices available to language users at the word, sentence, and discourse levels. Users are not seen as making these choices in a vacuum. Rather, choices about what is appropriate or useful are shaped by the user's knowledge of patterns in the immediate situational and cultural contexts.

According to Halliday, the situational context of any given text—oral or written— involves three dimensions that together constitute the *register* of the text. These variables and the interactions between them are reflected in the lexical and grammatical choices made by the participants. The first dimension, *field*, refers to the “what” of the text: what is happening, what kind of social activity is taking place, or simply what the text is about.

Taking for example a hypothetical analysis of a child's persuasive text about whether children should be allowed to choose their own pets, the text would likely feature language related to animals, the home, and perhaps responsibility; these would constitute the field. The second, *tenor*, concerns the "who": the relationship between writer and reader or speaker and listener, including expectations for behavior based on the nature of that relationship (e.g., cultural expectations regarding differences in status). If the text took the form of a letter to the child's parents, the student's lexicogrammatical choices would likely take a less formal tone, evidencing a close relationship. If the text were more academic in nature, the writer would want to sound more authoritative and objective, and to that end would likely use fewer personal pronouns and appeals. The third register variable, *mode*, is about the role of language: what it is expected to accomplish and how it is supposed to be organized. Both kinds of persuasive texts in our example would reflect a written mode, in which information is packaged and organized differently than it would be in speech.

Each of the three register variables is realized in language by a corresponding functional dimension or *metafunction*. Respectively, the *ideational*, *interpersonal*, and *textual* metafunctions work together to bring meaning to the text by construing experience, enacting social relationships, and organizing discourse (Martin & Rose, 2008). (For a detailed discussion of this study's use of linguistic constructs relevant to each metafunction, see Chapters 4 and 5.) Register variables and metafunctions operate simultaneously as social context informs and is realized by the text. Students (and most language users) are generally not making conscious decisions when they make choices related to these dimensions of language. Those who are steeped in the language and

culture in question can often present effective patterns of language without conscious attention to this, but this is not always the case for those who are not. A key strength of SFL-oriented pedagogy is that it offers a framework for making these patterns visible and accessible to learners. A starting point for this work is the concept of *genre*.

Treatment of genre. SF linguists define *genre* as a “staged, goal-oriented social process” (Martin & Rose, 2003, p. 7)—an evolving but fairly consistent, culturally-shaped pattern of expression or action. Considerable attention has been paid to building functional descriptions of the school genres that children must master to gain entry into dominant forms of discourse (e.g., Christie & Derewianka, 2008; Martin, 1989). There have been a number of studies of the use of functional, genre-based pedagogies with underserved student populations, including language minorities, in the Australian context (for a review, see Christie & Unsworth, 2005). In recent years, similar approaches (e.g., Brisk & Zisselsberger, 2010; Gebhard, Harman, & Seger, 2010; Schleppegrell & Oliveira, 2006) have become increasingly visible in the U.S. educational context.

As Gebhard and Harman (2011) observe in a recent review of genre theory in K-12 contexts in the United States, researchers and practitioners in other theoretical traditions approach genre somewhat differently in terms of the emphasis they place on language, learning, and social change. In addition to SFL, they identify English for Specific Purposes (ESP) and New Literacy Studies² (NLS) as genre-oriented perspectives that play, or could play, important roles in the K-12 landscape. Like SFL, ESP (e.g.,

² Gebhard and Harman explain their inclusion of New Literacy Studies rather than New Rhetoric Studies by saying that, while they share a number of features, “NLS approaches are more apt to involve conducting investigations related to the literacy practices of K-12 students in and out of school, especially in regard to how youth use emergent digital means of communication (e.g., gaming, blogging, texting). In addition, NLS approaches are not apt to involve conducting analyses of lexico-grammatical forms and rhetorical patterns of discourse” (p. 47).

Swales, 1990; Swales & Feak, 2004) uses linguistic tools to make relatively stable text structures accessible to native and non-native English-speaking learners. ESP's analysis of genres and the development of genre knowledge, however, has traditionally focused on academic and professional text types and for this reason is not often drawn upon by K-12 researchers or practitioners, though Gebhard and Harman point out that ESP practices of cataloguing and analyzing the features characteristic of a particular genre could have valuable applications at the secondary level in particular. The present study does not overtly draw on an ESP perspective, but there are some parallels between ESP efforts to induct writers into academic or professional discourse and this study's investigation of initiating students into a type of writing that will be increasingly asked of them as they become "expert" students (though, of course, this study cannot speak to longitudinal development).

NLS, for its part, defines genre much more broadly, as social action. Using critical ethnographic methods, NLS scholars (e.g., Gee, 2004; Heath, 1983; New London Group, 1996) emphasize the roles that issues of power and identity take in shaping the construction and interpretations of text, with a special focus on broadening conceptions of literacy to include literate practices in in- and out-of-school contexts, including online settings. Some scholars who have taken up this approach (e.g., Luke, 1996) have been critical of both SFL and ESP for reproducing, rather than interrogating, existing power structures through their emphasis on teaching established text types. Gebhard and Harman note that "this perspective of genre and genre teaching as a site of social struggle and as inherently multivoiced and hybrid makes some NLS scholars highly skeptical

about the merits and even feasibility of teaching genres to students as if they were fixed and stable” (p. 51).

While questions of power and agency are crucial in studying learners and in helping learners study texts, there are instances in which the rejection of formal approaches to teaching about language and genre can itself be a form of disenfranchisement. Contrary to the above critique asserting a fixed view of genre, SFL views offer an extensive framework for analyzing how genres shift according to contextual variables. Nevertheless, if SFL analysis of writing in schools has yielded descriptions of relatively stable school-based forms, it seems entirely appropriate to use these descriptions—adapted to local contexts (Schleppegrell et al. 2013)—as a starting point for helping students explore the textual practices surrounding them. I contend that providing children, especially those that do not come from mainstream cultural or linguistic backgrounds, with the linguistic tools that allow them to access and actively participate in dominant forms of discourse is fundamental to their empowerment. As Macken-Horarik (1998) noted, the concept of critical literacy presupposes certain linguistic proficiencies, without which it is difficult, if not impossible, to meaningfully engage with texts. “It is not fair,” she wrote, “to invite our students to critique texts before they have learned to analyze them and still less fair to those who cannot yet even process their meanings” (p. 78). Seen through this lens, SFL genre theory can be complementary to the NLS focus on interrogating traditional power structures and expanding conceptions of literacy. The ethnographic methods commonly used by NLS scholars make a strong case for why this work needs to be done. SFL adds to the mix a

“how”—a set of concrete pedagogical tools for helping teachers and students engage in deep exploration of texts and the way language choices shape the meanings within them.

The pedagogical approach presented in this study, in which children are invited to closely analyze and discuss in detail how meaning is constructed in one instantiation of argumentation, is one step in a long learning trajectory. Participants in the study have already encountered other persuasive genres, and, though the argument genre presented here attempts to represent structural, linguistic, and interpersonal elements of argumentation that will be valued in schooling contexts, they will certainly be presented with argument genres that rearrange and place differing levels of emphasis on these elements as they progress through school, and will continue to engage with other forms of argumentation in their out-of-school lives. No claims are made here, then, regarding the primacy of this version of argument, except as regards its potential to develop skills in logical argumentation that will become increasingly useful in their lives as students.

SFL and writing development. SFL has provided a useful framework for analyzing and describing the writing of students across a broad range of genres. One seminal work in this area is Christie and Derewianka’s (2008) analysis of approximately 2,000 student texts generated in primary and secondary English, history, and science classes collected over two decades of research in the Australian context (see also Christie, 2012). The authors describe the lexicogrammatical resources students of different ages use to realize genre-specific meanings across a wide variety of text types and in different content areas. Whereas much foundational work in writing development has focused primarily on sentence-level measures (e.g., Hunt, 1977; Loban, 1976; Perera, 1984) a major contribution of this work is to provide a functional description of writing

development at the level of the text. In addition, while use of linguistic resources varies by genre, the authors were able to synthesize findings from across text types to draw conclusions about how the use of specific language resources typically develops from one age group to another. These findings provide the basis for references in the present study to “developmentally appropriate” use of language features at the clause level. The present study expands on Christie and Derewianka’s work by adding text-level analysis of argumentative writing in two different age groups at the primary level; while they did analyze related texts written by adolescent students, their research does not include texts in this genre written by young children.

SFL studies of argumentative/persuasive writing. Only two studies of SFL applied to instruction in argumentation in K-12 contexts have been published recently. In the first, Gebhard, Harman, and Seger (2007) reported on an approach to teaching the academic language of persuasion to a class of fifth graders, many of whom were English learners. The researchers used a case study design to document the teacher’s instructional decisions as she planned and led her students through a unit built around a letter-writing campaign in which students communicated their dissatisfaction with not having recess to their principal. The teacher drew on an SFL genre perspective to teach about the social purposes of argumentation; its structural elements or *stages* (e.g., thesis, supporting arguments, counter-arguments, and restatement of thesis); and its associated language features (e.g., modality, connectors such as *therefore* and *although*, *if/then* syntactic structures). Although this SFL approach did not draw on Toulmin’s model, the stages of the genre as taught by the teacher aligned closely with Toulmin’s basic structural elements. Students also read and took notes on a brief literature review regarding the

importance of recess in elementary school so that they could draw on this information in building their case. Included and analyzed in the study were the free write, notes, first draft, and final draft produced by an English-learning focal student, allowing for a more detailed examination of individual response to instruction than the larger experimental or quasi-experimental studies afford. The study showed how the student appropriated and transformed the linguistic and organizational features presented in instruction for her own persuasive texts, further demonstrating the potential of elementary-aged children to make meaningful use of explicit instruction in argumentation.

Schulze (2011) also used SFL to structure a fifth grade unit on “persuasive arguments” (p.130). As with the study described above, the school in which Schulze carried out his action research project was an urban school with many cultural and linguistic minority students; 40% of the school’s students were native Spanish speakers and of those, approximately 80% were originally from outside of the United States. The persuasive unit took place in a pull-out class designed to support the academic English development of six “early emergent” (p. 135) students, one of whom served as a focal student whose work was at the center of the study’s discussion. Schulze led his students through a recursive genre instruction cycle (known in SFL as the teaching-learning cycle; for a detailed discussion, see Chapter 3) including reading and discussion of a number of sample texts; collaborative writing of a persuasive text; multiple drafts of student texts; and special attention to the language used to approximate an academic register, strengthen or soften tone, and transition effectively from one part of the text to the next. Although Schulze provided an otherwise detailed description of his instruction, he did not articulate which stages were included in the persuasive genre(s) he taught, nor did he

present the prompt to which students responded. The focal student's first draft contained a number of features characteristic of persuasive letters, while subsequent drafts adopt features of academic argument. Based on the student texts presented, it is clear that the topic was the 2008 presidential elections; later drafts indicate that second graders at the school were the intended audience.

After describing the instructional cycle, Schulze presented a register analysis in which he discussed the features of these three drafts in great detail. His SFL register analysis examined the variables of *field*, *tenor* (including mood, modality, and appraisal), and *mode* (including analysis of Theme/Rheme, conjunction, and repetition) at the clause level. Through this analysis he demonstrated that with progressive drafts field was refined, tenor became more distant and authoritative, and there was an increase in the use of conjunctions to link the focal student's claim with supporting details (for example, a fourfold increase between first and final drafts in the student's use of the connector "because" to link and elaborate on ideas). These findings were illustrated with the specific examples of the focal student's writing, but were representative of the changes Schulze observed in other students' work, again showing how a dual focus on language and content can improve children's academic writing.

The studies carried out by Gebhard, Harman, and Seger (2007) and Schulze (2011) exemplify the rich detail that is one of the primary strengths of SFL approaches to instruction and analysis. Instruction in both cases was carefully planned to give students the opportunity for deep exploration of the target genre. Children examined model texts, discussed purposes for writing, and looked closely at how language was used to present and structure ideas; they also read and discussed other texts to enhance their knowledge

of the issues about which they would be writing. While some of the non-SFL interventions reported on above also integrated some of the same features into their instructional designs, a focus on how to use language—especially written language—to achieve social/rhetorical goals beyond simple “sentence starters” is noticeably absent. SFL also provides a framework for fine-grained analysis of the written products of instruction, allowing researchers to identify the precise ways in which children’s writing changed. In contrast, the holistic scoring and numerical scales characteristic of most other studies of children’s argumentation can confirm *that* the quality of student writing has improved, but not *how* it has done so. In order to strengthen instruction by targeting it to students’ specific needs, a robust theory capable of pinpointing what their linguistic needs is called for. To that end, SFL is employed in this study’s instructional design and used in data analysis to characterize student performance.

While the two SFL studies discussed here use methods similar to this study’s to describe how explicit discussion of genre can support elementary students’ learning/writing, they also differ in some important ways. For instance, both report on the writing of a single class of fifth graders, leaving open the question of the extent to which similar instructional methods would be effective for younger students. Moreover, each article discusses close analysis of the work of just one focal student. As a practical matter this decision makes perfect sense, given the level of detail included in the respective analyses. However, reporting results in this way involves a tradeoff: readers learn a great deal about the performance of one representative student, but not about the range of performance in the classroom. The present study addresses these questions by examining argumentation in second and fourth grade, thereby generating information about

performance at these younger grades and providing some basis for comparison of the characteristics of writing at each level. Furthermore, this study takes whole classes, language proficiency subgroups, and academic proficiency subgroups as units of analysis in addition to describing the work of several focal students in each grade. By looking at the problem from several different angles, it contributes additional nuance to the discussion of the kind of argumentative writing a single student, or even students at a particular grade level, can be capable of if learning is properly supported.

Key constructs for the present study. The present study took place in the context of a larger research project that provided a group of elementary school teachers with ongoing professional development in an SFL-based instructional approach called Functional Grammar (FG). Over the course of the year, teachers learned about four key genres: *recount*, *character analysis*, *description/explanation*, and *argumentation*. Genres in FG are described in terms of their social purposes, structural elements (hereafter referred to as *stages*), and language features typically used to build meaning. Stages and language features are presented in terms of how they work to help writers achieve their goals. For example, teachers may use a model text to discuss a genre's stages, foregrounding the function of each: What important information is being communicated to the reader at this stage in the text, and how? Similarly, functional labeling is used to describe grammatical constituents. For example, *processes*, *participants*, and *circumstances* are FG constructs that align roughly with verb phrases; noun phrases; and adverbs and prepositional phrases, but focus students' attention on the larger units of meaning that allow them to explore what a text is about, how it is organized, or how judgment/evaluation is expressed (Schleppegrell, 2004). This metalanguage allows

students to engage with the kinds of meanings relevant to their genre-specific purposes. The following table describes a subset of the metalanguage that was used in this study to support argumentative writing in which students attempt to convince a reader of the merits of their position by providing supporting evidence and anticipating potential counterarguments:

FG Metalanguage	Function	Use in argumentative writing
Modality	Expresses degrees of possibility or obligation/necessity; opens or closes dialogue to other opinions	Students experiment with altering modality to make their own recommendations strong (<i>we need to</i>) or weaker (<i>we could</i>) or, at more advanced levels, to distance themselves from counterarguments (<i>Some people think..., I'm sure</i>).
Connector	Links ideas between or within sentences (conjunctions or text/sentence connectives)	Contrastive connectors like <i>however</i> or <i>on the other hand</i> are used to introduce a student's rebuttal to an anticipated counterargument; other connectors organize text as a whole.

Table 2.1. Metalanguage used to support argumentative writing.

The strength of an FG approach is perhaps most evident in the far-right column of the table, which shows how the constructs were deployed in the classroom. The metalanguage does not assume the form of static, declarative knowledge, but is actively used as a tool to deconstruct and explore authors' language choices and the choices students themselves can make as they construct their own arguments. Other features of argumentative writing (e.g., generalized participants; mainly timeless present tense; use of nominalization to present argument more objectively) were observable in model texts but were not the subjects of explicit instruction. Chapters 4 and 5 provide detailed accounts of the argument genre descriptions and instructional sequences in second and fourth grade, respectively.

In the next chapter, I provide a description of the research context, the research questions guiding my study, and the methods used to collect and analyze data in pursuit of the goals outlined therein.

CHAPTER 3

Methods

Overview of Larger Study

This study was carried out within the context of a larger research project. The Language and Meaning Project, focused on the design and implementation of FG-based professional development for teachers of ELLs, officially began in 2010, though Functional Grammar (FG) work with teachers in the participating school district, Oak Grove³, had been taking place since 2006. The project's participants during the 2011-2012 school year were second- to fifth-grade teachers and literacy coaches representing six schools, all of which were particularly well-suited to a study of ELL writing: Approximately 90% of their students had a first language other than English (primarily Arabic). The urban district of which they are a part has the highest concentration of ELLs in the state of Michigan. The schools share several key characteristics with other U.S. communities where there is a high concentration of ELLs, including a large number of students from low-SES backgrounds, frequent transfers into and out of the schools/district, and limited opportunities for some students to learn English outside of the classroom.

The purpose of the Language and Meaning Project is to provide teachers with an SFL-informed pedagogical approach for exploring and talking about the language of texts in their classrooms. Participants in the project learn about functional metalanguage, and

³ A pseudonym.

how it can be used to explore texts and engage students in meaningful talk about their reading and writing. They also learn about the FG approach to carefully scaffolding writing instruction, including explicit discussion of a genre's social purposes, component stages and their functions, and language features important to building the kinds of meaning that help writers achieve the goals of the genre. While the FG approach aims to improve academic outcomes for ELLs by being explicit about the way language works in academic settings, there is a basic level of proficiency needed to engage meaningfully in the instructional environment under study. Newcomers who have not yet attained this minimum level of English proficiency may, however, participate in FG lessons at their teachers' discretion. In the present study, for example, while newcomers' written work was excluded from data collection, they were not excluded from instruction: These students participated in a modified version of the curriculum taught by an Arabic-speaking resource teacher and on some occasions joined their classmates for discussion and other activities.

Language and Meaning participants in 2011-2012 attended a week-long preparatory summer workshop and participated in three additional day-long workshops over the course of the school year, each focused on teaching a different writing genre. Before starting the argumentative writing unit in April, teachers and students had experienced three genre-based units—Recount, Character Analysis, and Description and Explanation—and become regular users of the FG metalanguage (e.g., stages, processes, connectors, etc.). After each workshop, members of the Language and Meaning research team observed the participating teachers' implementation of the FG units. These observations, along with teachers' direct feedback, were used to inform the development

of the professional development and classroom instructional materials for the subsequent unit(s).

Overview of Dissertation Study

In working to fine-tune the Functional Grammar pedagogy, members of the Language and Meaning Project frequently worked with classroom observation data. These data helped the team understand, broadly speaking, which aspects of the approach made the FG constructs accessible to teachers (and, in turn, their students) and which elements proved challenging. Given the time constraints under which curriculum planning had to take place, writing data were given less attention. Having spent time developing plans and materials, delivering professional development, and observing in classrooms as a member of the research team, I became interested in what else the writing could tell us. Children's writing is a rich source of data on their experiences: if looked at in enough detail, it reveals not only what the writers consciously chose to express, but also shines a light on how well they were supported to achieve instructional goals and on specific strengths and needs that we might otherwise not be aware of.

With this in mind, I decided to closely examine the writing produced in the unit on argumentation. Table 3.1 provides an overview of this study's research questions, data sources, and coding/analysis strategies. My project uses a case study research design, taking as its cases a) two intact classrooms and b) 10-11 focal students within each of those classrooms, two to three of whom provide illustrative examples of student work and interview responses at each grade level. Cases were constructed using qualitative methods for collecting observational data and artifacts as well as interviews; these methods are described in detail in the following sections of this chapter.

Research question	Data sources	Coding/Analysis strategy
<p>What were the features of students' argumentative writing at the second and fourth grade level when supported by a Functional Grammar pedagogy? How do the features of students' arguments differ by academic achievement or language proficiency levels?</p>	<p>Final drafts (class sets for 2nd and 4th grades; 25 and 26 arguments, respectively)</p>	<ul style="list-style-type: none"> • Code for presence and effectiveness of genre stages, establishing range of performance for grades and subgroups • Analyze clause-level features associated with expectations for the genre
<p>What were the Functional Grammar-informed argument curriculum and other features of the instructional context in which English learners' written argumentation in second and fourth grade was observed? How was this reflected in the features of student writing? Does scoring of pre- and post-instruction writing using a primary traits rubric show that students' arguments improved after participating in instruction?</p>	<ul style="list-style-type: none"> • Classroom observation records (video, audio, and fieldnotes) • Final drafts • Pre- and post-instruction writing • Formal interviews with focal students (reflections on writing) 	<ul style="list-style-type: none"> • Describe how teachers used the FG curriculum to support students in preparing to write • Analyze student writing to characterize ways in which it was shaped by classroom discussion, scaffolding, etc. • Have pre/post writing evaluated by independent raters to further determine impact of instruction • Code focal student interviews to evaluate genre knowledge, student perceptions of pedagogy
<p>What relationship, if any, exists between students' performance as measured by linguistic analysis of the structural components of their arguments and the scores they receive on a traditional primary trait assessment? How do these approaches to assessment differ in terms of the potential of the feedback they generate to inform development of argumentation skills?</p>	<p>Non-participant teachers' primary trait scoring of essays</p>	<ul style="list-style-type: none"> • Compare scores with writing performance across dimensions of genre knowledge

Table 3.1. Overview of research questions, data sources, and analysis strategies

Sampling

Data for the classroom cases were collected from a cross-sectional sample consisting of two classes: one second-grade (n=25) and one fourth-grade class (n=27, of

which 26 essays were ultimately included in data analysis), at separate schools. While it would have been ideal to select teachers from the same school, participating schools generally selected two teachers each from a limited range of grade levels to be involved with the larger project. As a result, there were no schools in which both second- and fourth-grade teachers participated in FG professional development. However, the Oak Grove district had instituted a framework for literacy instruction across schools and carefully monitored its implementation; this policy, along with shared reading programs and similar demographics⁴, meant that the schools had much in common.

Because the questions under investigation depend on a rich instructional environment, the classes selected were those of teachers who had been identified as successful implementers of FG instruction over the course of the year. Successful implementation involves not just transmitting the targeted concepts accurately, but providing appropriate levels of scaffolding and productively managing opportunities for students to interact with one another as they learn. (Scaffolding is used here in its expanded sense, referring to mediation of understanding through the use of both coaching behaviors and cultural artifacts; see Kamberelis & Bovino, 1999, for discussion.) Choosing to pilot the unit in classrooms of high implementers meant working with students who were likely to have a solid grounding in the relevant FG constructs, thus limiting a potential impediment to success.

⁴ According to data provided by the state, in 2011-12 ELLs made up 70.25% of the second grade school's population and 69.84% of the fourth grade school's population. 93.55% and 92.26% of students at the second and fourth grade schools, respectively, were economically disadvantaged. The second grade school was 100% white and the fourth grade school was 98.39% white, following the U.S. Census Bureau's definition of "white" as including people of European, Middle Eastern, or Northern African origin; the overwhelming majority of students at both schools were of Middle Eastern origin.

Additional sampling considerations applied to the selection of focal students within each class. 11 second graders and 10 fourth graders were chosen as focal students. These students were selected to represent a range of language proficiency and academic achievement levels. Before beginning data collection, language proficiency profiles were built for each class using standardized testing information in consultation with the teachers. In the participating schools, all children who have a native or primary language other than English are assessed using Michigan's English Language Proficiency Assessment (ELPA). ELPA measures proficiency in reading, writing, listening, and speaking, and reports five performance levels: Basic, Low Intermediate, High Intermediate, Proficient, and Advanced Proficient (see Figure 3.1). Students are tested in the fall and again in the spring to assess progress. Though the spring testing had been carried out shortly before data for the study were collected, those scores were not yet available. Accordingly, fall data were used in consultation with teachers, whose day-to-day experience with their students added valuable detail about current performance to the quantitative data provided by ELPA, as well as provided information about the children's home language, the amount of time they had been learning English, and other relevant details. Spring ELPA data were consulted once they became available and confirmed that the basic language proficiency profiles of the classrooms had not changed.

The spring data showed all second-grade participants in the study testing as English learners, ranging from levels 5-1 (Basic to Advanced Proficient⁵) on the ELPA. Focal students at this grade tested at levels 5-2 (see Table 3.2). In the fourth-grade class, just under half of the class had tested out of ELL services, with the remaining students

⁵ Only one second grader tested at the Advanced Proficient level; spring test scores for the rest of the class are roughly evenly divided between the remaining four proficiency levels.

classified at ELPA levels 3-1. These scores represent language learning progress for the fourth graders: all students for whom prior years' data were available had started kindergarten classified as ELLs. Table 3.3 shows the ELL status/proficiency levels of the fourth-grade focal students.

Basic	(B) Level 5
Low Intermediate	(LI) Level 4
High Intermediate	(HI) Level 3
Proficient	(P) Level 2
Advanced Proficient	(AP) Level 1

Fig. 3.1. Michigan's English Language Proficiency Assessment (ELPA) levels

Teachers were also asked to evaluate individual students' general academic performance as a dimension distinct from language ability so that focal students could be chosen to represent a range of academic achievement as well as language proficiency levels. Both classroom teachers chose to use the students' reading level as a proxy for their overall academic performance. In the second grade, the three possible reading levels were *At/Above Grade Level*, *Some Risk*, and *At Risk*. The fourth grade teacher used four reading levels: *Above*, *On*, *Approaching*, or *Below Grade Level*. Reading levels tended to roughly correspond with language proficiency levels, though this was not always the case. Both teachers recommended specific students within and across reading levels based on a variety of personal characteristics, including levels of classroom participation and enthusiasm, creativity, perseverance, and a general impression that the student would be "interesting to follow". These recommendations were combined with my own in-class observations of and interactions with students to make a final selection of focal students at each grade level (see Tables 3.2 and 3.3 below).

Reading Level	Pseudonym	ELPA Level
At/Above Grade Level		
	Amina	2
	Nadia	2
	Fatima	2
	Kamil	2
	Sabah	5
Some Risk		
	Hassan	4
	Basinah	4
	Ali	3
At Risk		
	Zaineb	5
	Samir	4
	Leila	3
	Salma	(no data available)

Table 3.2. Second-grade focal students

Reading Level	Pseudonym	ELPA Level
On Level		
	Isa	Monitored
	Omar	Monitored
	Jamal	Monitored
	Hadiya	Monitored
	Sanaa	Monitored
Approaching Level		
	Amir	3
	Khadijah	3
Below Level		
	Samirah	1
	Iman	3
	Maryam	3

Table 3.3. Fourth-grade focal students

Data Collection

Classroom data collection for this study took place over a roughly one-week period in late April 2012. The broader unit of study for each classroom was

argumentative writing lesson sequence, which consisted of five approximately 45-75 minute lessons (60 minutes on average) designed to support students' knowledge about the particular topic they were exploring as well as the relevant writing and reasoning skills that needed to be developed for the argument genre. Lessons in the second grade were primarily taught by one of the larger project's Principal Investigators, and the fourth-grade lessons were taught by an advanced graduate student who had several years of experience working as a research assistant on the project. I was present for all lessons. I co-taught a number of lessons in the second grade and interacted with students as they worked; in the fourth grade I offered support while students worked in groups/pairs or independently and taught the final lesson of the unit.

When not teaching, I observed instruction and prepared detailed fieldnotes. Video and audio data provided a thorough record of simultaneously occurring classroom events that could be consulted after the lesson had concluded. Several video cameras were used to capture interaction from different angles and in different parts of the classrooms (e.g., from the front of the room looking out at the students; from the students' perspective looking at the teacher and scaffolds; and at the sides of the room where a particular group or pair of students was seated). Separate audio recorders were used to record conversation during pair or small group work. These recorders were placed on the table in the center of the selected group or pair—which always included at least one focal student—and produced two to four separate audio files per lesson, depending on the configuration of students in a particular activity. These data provided a detailed picture of how the classroom environments supported ELLs' learning and allowed for later identification of key teacher-student(s) or student-student(s) interactions, with particular attention to

conversations that serve to expand or redefine students' working definitions of a construct.

I conducted formal interviews with 11 students at each grade level on the day they finished writing their arguments⁶. These interviews added a measure of students' metalinguistic awareness and rounded out observational data with students' own perspectives. The formal interview protocol (see Appendix A) consisted of 16 questions prompting focal students to reflect on what they had enjoyed or found challenging about the unit; the ideas and artifacts or activities they drew on while writing; the position they took and why; and whether they believed their arguments were effective. In addition, students were asked to describe and explain their work in terms of a) the overall text structure, stages, and purposes for writing and b) particular linguistic features by which the former are realized, thereby eliciting explicit knowledge of both macro- and micro-level genre features. Interviews lasted seven to eight minutes on average but were expanded when students needed additional prompting or discussion in order to answer questions, or contracted to avoid discouraging students when it was clear that, even with additional prompting, they were unable to answer.

The students' written arguments were the principal data source for investigation of the features of ELLs' argumentative writing across grades. Complete arguments were collected for 25 second graders and 27 fourth graders. All students also completed both pre- and post-instruction argumentative writing tasks, using the same prompt on both occasions. In total, then, written data included 75 second-grade writing samples and 81

⁶ Two second grade students were interviewed on the next school day following a weekend break; extra time was taken for discussion in these cases to maximize students' recall of the prior week's events. These discussions followed the protocol with the addition of the question "What did we learn last week?" and the shifting of the question "What was your favorite part of what we did?" to the beginning of the interviews.

samples from the fourth grade. After the conclusion of the unit, these writing samples were sent to independent raters for scoring using a primary trait rubric.

Data Analysis

Several types of analysis were used to explore the writing data in this study. This section presents methods for clause-level analysis and independent scoring of student texts. Detailed descriptions of methods for evaluating the effectiveness of arguments at the stage level are provided in Chapters 4 and 5.

Clause-level analysis. This study's approach to clause-level analysis was developed and refined in cooperation with an SFL expert. The process of defining the analytic approach began with discussion of the linguistic features of most interest given the goals of the study. This discussion led to the development of a draft analysis template that identified the features of the texts that the analysis would focus on and what that analysis would reveal. For example, *cohesion* was analyzed by looking at the ways the writers introduced and tracked participants in their texts, *ideas and content* were explored through transitivity analysis (process types), and *voice* was analyzed by examining use of modality, mood, and emphasis realized through capitalization, underlining, punctuation, and the like. We conducted independent pilot analyses of three arguments from second and fourth grade and further refined the template to focus on types of analysis that yielded the most useful results. This refined set of analyses, all of which I had been trained in my graduate education to perform, included: clause, T-unit, and embedded clause counts to facilitate comparison of clause complexity between groups by analyzing number of clauses per T-unit and embedded clauses per T-unit; participant tracking to explore issues related to reference; transitivity analysis to characterize how meaning was

built through choice of different types of *process* types (doing, saying, thinking/feeling, being); *connectors*, *circumstances*, and *marked themes* were tabulated and analyzed for the meanings they contributed in order to discuss students’ text organizational choices; and analysis of language choices related to mood and modality informed discussion of the *voice* projected in the students’ texts (see Table 3.4).

Analysis	Approach
Clause analysis	Count total numbers of clauses, T-units, and embedded clauses; calculate number of clauses per T-unit and number of embedded clauses per T-unit
Participant tracking	Identify reference chains (participants referred to more than once)
Transitivity and theme analysis	Render text into ranking clauses and parse clauses into participants, processes, circumstances, and connectors; identify process and connector types and tokens; identify marked themes
Modality and Mood	Identify modality usage and type (obligation, ability, likelihood); note instances of interrogative or imperative mood

Table 3.4. Approach to clause-level analysis

I subsequently analyzed this set of features in all student arguments. Where questions arose in the course of analysis, I consulted with the SFL expert and we came to consensus about how specific issues in this particular data set would be treated. These included such issues as the definition of clause and/or T-unit boundaries in ambiguous cases and proper analysis of infelicitous language. One example, from the second grade data, is the T-unit “They won’t know how to catch fish and lots of space like 20 Oak Groves,” in which “lots of space like 20 Oak Groves” should have been a clause with a separate verb but is not (e.g.: “They won’t know how to catch fish and they won’t have lots of space like 20 Oak Groves.”). Because the student did not construct this as two clauses, our decision was to treat the noun group “lots of space like 20 Oak Groves” as part of the embedded clause beginning with “how to catch fish.” This passage was analyzed, then, as one T-unit with an embedded clause.

Another example of an analytic question that was resolved in discussion comes from the fourth grade data and concerns the abbreviation of the possible claims, a

practice that began in classroom discussion and later made its way into student writing. The claim “Meat ants are a good option but need to be studied more” was shortened to “good, study more”; similarly, the claim “Meat ants are a bad solution and should not be used” became “bad, don’t use.” One student wrote “I think they should not use bad don’t use because good study more is better.” Our decision in this case and others like it was to treat “bad don’t use” and “good study more” as participants (noun groups), as they function as noun groups in the texts (in the example sentence, we can see that these can be replaced with demonstratives or pronouns, as in *I think they should not use that because this is better*).

In total, we consulted on particular issues that arose in 27 student arguments, or 53% of the total sample. Of these, 19 came from the fourth grade, which I analyzed first; we discussed eight of the shorter and less complex second grade arguments. As we discussed each instance and I received guidance on the appropriate way to code that token, I made notes about the coding decision and revisited data I had already coded to ensure that the principle adopted for the particular case was also informing analysis of the rest of the corpus. Detailed discussion of clause-level analysis in grades two and four are presented in Chapters 4 and 5, respectively.

Independent scoring. While SFL analysis allowed for a fine-grained assessment of student writing, it is not widely used by teachers or designers of high-stakes tests. Recruiting local teachers to rate the second and fourth graders’ arguments using an ecologically valid scoring instrument provided insight into how Oak Grove teachers would assess students’ pre- and post-instruction writing and established a basis for

comparison between SFL analysis and a traditional primary trait assessment of the arguments students produced during the FG unit.

Identification of assessment. The primary trait assessment chosen for comparison was an adapted from the popular 6+1 Traits Writing Model used in the district (Culham, 2005). Under this model, writing is scored with rubrics emphasizing six key constructs: Ideas, Organization, Voice, Word Choice, Sentence Fluency, and Conventions, with a separate focus on Presentation (the eponymous “+1”). The Oak Grove district had been using the 6+1 Trait Writing Model for its classroom- and district-level writing instruction and assessment for several years, and had used the model to develop its own genre-specific rubric for grading arguments, or “opinion” writing (see Appendix B for the second- and fourth-grade versions; the rubric includes the six main traits but not Presentation). All Oak Grove teachers had used the rubric to score student responses to a district-level writing prompt in Spring 2012. This rubric, then, had the advantage of being familiar to teachers across the district—the population from which independent raters would be drawn—as well as being based on a widely-used paradigm for teaching and assessing writing.

Selection of independent raters. There were two key considerations in selecting teachers to serve as independent raters. First, teachers needed to be familiar with the study’s student population such that their expectations would be generally aligned with the range of performance they would be assessing. In other words, familiarity with the student population was expected to produce scores that would not be either artificially inflated or suppressed, as they might be if the raters were accustomed to working with populations that performed significantly better or worse than the students in the sample.

Second, while some exposure to Functional Grammar/SFL was acceptable—and would have been difficult to avoid within the Oak Grove district—it was established that raters should not be familiar with the larger research project’s current curriculum, as this, too, may have shaped their approach to scoring the writing samples in a way that reflected values other than those emphasized in traditional assessment. (For example, the FG approach puts less emphasis on conventions than does typical writing instruction, and takes a different approach to organization.)

To find teachers meeting these criteria, I selected schools within the district whose student populations were similar to the two schools where data were collected in terms of ELL concentration and socioeconomic status. Literacy coaches at these schools then assisted in identifying qualifying teachers interested in participating in the study. The three second-grade raters had been classroom teachers for 6-20 years, and had taught second grade for an average of six years each. The three fourth-grade teachers had taught from 11-15 years, with an average of 4.7 years in fourth grade. Four of the six independent raters reported that they were fluent in Arabic. All six had recently used the district’s Opinion rubric to score student writing at their respective grade levels and were experienced in the use of 6+1 Traits rubrics more broadly.

Scoring procedure. To prepare student writing samples for independent scoring, each piece was scanned and all identifying information was removed from the electronic files. Samples for all prompts—the pre-test, post-test, and final argument (the argument produced over the course of the unit) at both second and fourth grade—were pooled (n=156) and assigned random numbers from 1-156. Samples were then separated by grade and assigned to independent raters. Raters received a roughly equal number of

samples for each of the prompts at their grade level—a mix of main arguments and pre- and post- instruction tasks, as independent scoring for all tasks was done concurrently—and assignments overlapped such that each sample would initially be reviewed by two raters. These pairings were made on a rotating basis so that samples were not always reviewed by the same two raters. Raters were unaware of the order in which the writing had been produced, and were not aware that the samples they were evaluating included responses to a pre- and post-instruction writing task.

Communication with independent raters took place via e-mail so that teachers could complete the scoring at their convenience. Their initial instructions included a description of the scoring task; relevant background on the prompts students wrote to and summaries of the source material for the main argument and the pre-/post-instruction task; a copy of the rubric (see Appendix B); and a scoring sheet. The scoring sheet consisted of separate columns for each of the six dimensions being measured (no composite score) and space for raters to comment on any aspect of the text they wanted to, e.g., issues that made the text particularly weak or strong.

Once all raters had submitted their scoring sheets, initial scores were compiled and exact and adjacent agreement percentages were calculated. Essays for which there was a discrepancy of more than one point on any of the dimensions were flagged for a second “tie-breaker” round of scoring. In other words, if Rater 1 gave an essay a score of 4 for Voice and Rater 2 gave a score of 5, those adjacent scores were considered sufficient for that piece. However, if Rater 1 gave a score of 3, yielding a 2-point difference in scores, the essay was referred to the third rater (who had not reviewed it in

the first round of scoring). The final score for each essay was determined by averaging both or all raters' scores⁷.

Guiding Principles for Curriculum Design

In designing the argument curriculum for second and fourth grades, several guiding principles drove decisions both large and small. These principles included providing explicit, well-structured support to mentor learners into a new genre; keeping content cognitively manageable; and providing frequent opportunities for meaningful interaction. In the following section an explanation of these general principles is provided; for a detailed account of how instruction unfolded at each grade level, see Chapters Four and Five.

The argument curriculum followed a similar structure at both grade levels. This structure was informed in part by the teaching-learning cycle (Rothery, 1996). The teaching-learning cycle, developed by SF linguists working with teachers in the Australian context, is a pedagogical framework for exploring new genres with students. The version of the cycle drawn upon in designing the argument curriculum features three main stages: deconstruction, joint construction, and independent construction (see Fig. 3.2). At the deconstruction stage, students read and discuss model texts in the target genre. The joint construction stage involves teacher and students collaborating in a guided construction of a new piece in that genre. Subsequently, as its name would suggest, the independent construction stage provides students the opportunity to generate their own writing in the target genre. It is important to note that the setting of social

⁷ Discarding discrepant scores in cases where there was still some disagreement after the tie-breaker round was considered as an alternative but ultimately rejected. Including all three scores increases validity by more accurately reflecting the range of opinions provided by independent raters; conversely, discarding outliers may have obscured relevant differences in the way raters viewed students' writing.

context and building of field knowledge are synchronous and ongoing throughout the teaching-learning cycle. In other words, while there may be a certain amount of up-front study of the content, students do not “finish” learning about the subject matter and move on to learning about the genre, because discussion of one is integrated into discussion of the other.

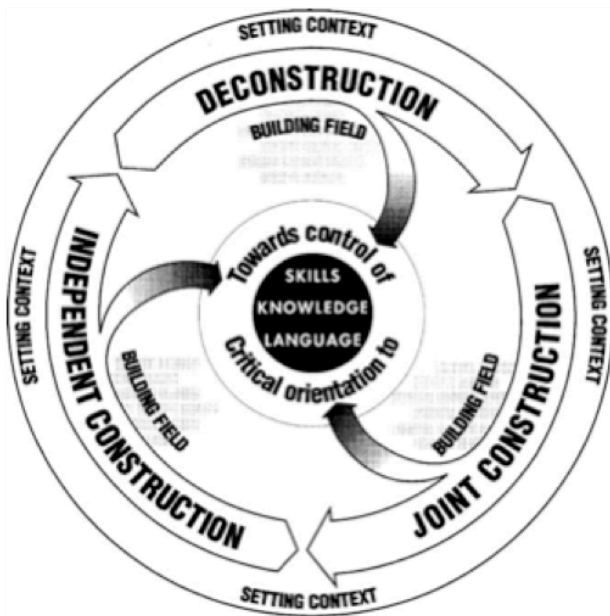


Figure 3.2. The teaching-learning cycle (from Martin, 2009, p. 16).

Another guiding principle in the development of the argument curriculum was that of cognitive manageability. In planning the lessons, the research team took manageability into account in several ways. First, because students would be working with informational text, it was seen as advantageous to build on students’ prior knowledge where possible. Informational text has features distinct from narrative, and elementary-level students are exposed to the former much less frequently than they are the latter (Crowhurst, 1980; Duke, 2000; Freedman & Pringle, 1988; Jeong, Gaffney, & Choi, 2010). Despite these challenges, students would have to glean from their reading

understanding of a number of related concepts in order to write their arguments. To that end, argument topics were chosen for each grade that tied into a theme that had emerged in the Functional Grammar unit that students had just completed: that of human impact on various ecosystems.

This decision had consequences for our treatment of audience, purpose, and authenticity, issues are known to affect students' writing and reading performance (see, e.g., Graham, McKeown, Kiuahara, & Harris, 2012; Midgette, Haria, & MacArthur, 2008; Purcell-Gates, Duke, & Martineau, 1997). The instructional design emphasized the overall social purpose of argumentation, and especially the functional purposes of its stages, throughout the argument unit, but choosing a highly authentic topic—one in which students might feel personally invested—proved to be a challenge. A number of possible issues specific to the local context were discussed, but we were unable to identify any that were locally relevant; provided sufficient opportunities to argue for multiple positions; and connected with the theme of human impact in a satisfactory way. The issues ultimately chosen, while not germane to Oak Grove students' everyday lives, built strategically on the concept of the consequences of human interference on various ecosystems. As will be detailed in Chapters 4 and 5, features of the instructional design such as a debate-like activity in second grade and an attention-grabbing video in fourth grade were included to enhance student engagement with the respective topics.

The selection of issues removed from students' everyday contexts, combined with the theory of language guiding the instructional design, limited the range of plausible audiences for students' writing. Prior research has found that having an external audience specified improves writing quality (Cohen & Riel, 1989; Crowhurst & Piché, 1979), and

while these studies were undertaken with older students, they suggest that asking students to write for peer audiences in particular can strengthen aspects of their writing, and recent research with young children (e.g., Block, 2013) has also found this to be the case. At the same time, the target genres in the larger research project are intended to be school genres, and instruction focuses on features reflecting an academic register. Specifying a peer audience would produce a shift in the register variable of *tenor* (see Chapter 2). The language features typically involved in addressing a friend differ from those used when there is a more socially distant audience, as is generally the case with academic writing. Our intention was to provide support for the linguistic challenges associated with the latter, so we did not explicitly specify an audience other than the teachers. We did, however, draw informally on the idea of directing a message to other audiences to help students plan their writing in cases where it was necessary to consider the perspective of someone on the other side of the issue or provide sufficient background information for someone outside of the classroom (see Chapters 4 and 5 for specific examples).

Another consideration related to cognitive manageability was how to distribute the amount of new information presented across the unit. Because of the additional demands posed by working with informational text, and because of the level of challenge involved in learning about the argument genre itself, each of the five lessons in the unit had to be carefully balanced in terms of the amount of new material introduced. One way the amount of new material was managed was by attempting to keep text length appropriate. The informational texts providing most of the field knowledge for the unit were constructed by the research team using information gathered from various sources and edited so as to provide sufficient content to help students understand the problem

without overloading them with information they would not need. These texts were limited to one page, 14-point font, for the second graders and two pages, 12-point font, for fourth graders. No established measures of text complexity were calculated, but efforts were made to simplify the original text while preserving meaning and scientific accuracy where warranted (for example, breaking complex sentences into multiple sentences, eliminating irrelevant dependent clauses, and removing researchers' names and institutional affiliations).

Another strategy used to keep the amount of new content manageable was to focus on one stage or set of stages at a time. During the first lesson, students were presented with an overview of the genre, including its component stages and their purposes. Experience with argumentation in prior incarnations of the Functional Grammar curriculum had shown that asking students to think about all of these aspects at once could be overwhelming. With this in mind, instruction in subsequent lessons narrowed in to focus on the development of the language and logic particular to specific pieces of the genre: Description of Problem and Claim; Evidence and Reason; Counterargument and Evaluation of Counterargument. Accordingly, students were provided with a number of graphic organizers designed to help them think through and draft the different stages of their arguments (see Appendices C and D for second- and fourth-grade versions, respectively) rather than a single organizer laying them out in fixed order. Focusing on the stages separately, it should be noted, did not mean focusing on them in isolation: discussion of the stages was recursive and aimed to locate their construction in the broader context of the overall purpose(s) for writing the argument and the interrelatedness of all its component parts.

Finally, careful attention was paid to providing plentiful opportunities for students to interact verbally with teachers and with one another throughout the unit. High-quality conversation around reading and writing promotes academic language development and English learners in particular benefit greatly from meaningful interaction in the classroom (Ellis & Larsen-Freeman, 2006; Lindholm-Leary & Borsato, 2006). In the case of writing, classroom talk allows students to rehearse, clarify, and reformulate their ideas before committing them to print (Horowitz, 1987). Modeling the pedagogy on the teaching-learning cycle meant that reading, writing, and discussion were concurrent throughout the argument unit, and this interrelationship afforded a number of occasions in each lesson for focused conversations between partners or members of a small group. In addition, the instruction itself was designed to be interactive, giving students the chance to both share their knowledge and offer ideas about the dilemmas driving the arguments that in turn helped determine the course of instruction.

The next chapter discusses data collected in the second grade, beginning with a detailed account of how these principles were realized in classroom instruction.

CHAPTER 4

Second-Grade Writing

This chapter provides context for the production of second graders' arguments and presents the results of writing analysis. The chapter begins with a description of the research context and the day-by-day unfolding of lessons in the argument unit. This description is followed by presentation of scores assigned to pre- and post-instruction writing by independent raters, indicating that most second graders wrote more highly-rated arguments after participating in the argument unit. Next, I report on clause- and stage-level analysis of student arguments for the second grade class as a whole as well as academic achievement and language proficiency subgroups, using examples to illustrate the range of performance observed in writing at this grade level. This characterization of student writing yields a number of pedagogical recommendations tailored to the specific aspects of argumentation that proved challenging for second graders in this study, including findings regarding the surprising ways in which source texts shaped children's arguments. Analysis of student writing also resulted in the development of a detailed framework for the evaluation of arguments; this framework is introduced in the latter part of this chapter. The cases of two focal students are used to show how the different analyses discussed throughout the chapter can come together to paint a detailed picture of the work of an individual child as well as to highlight differences in performance by students in the same grade. Finally, I present the results of independent raters' primary

trait scoring of the second grade writing, which support several key findings of my analysis and call attention to important differences in the values encoded in SFL-informed evaluation as compared to traditional assessments such as the primary trait rubric used in this study.

Research Context

The second-grade classroom chosen for piloting the argument unit was that of Ms. Beydoun, one of two second-grade teachers participating in the Language and Meaning Project during the 2011-2012 school year. Ms. Beydoun had a Bachelor's degree in Elementary Education with an emphasis in Math and Language Arts and had been teaching at the first, second, and/or third grade levels for a total of four years. She had a bilingual certification and was herself bilingual, having grown up speaking English and Arabic at home. She reported her fluency level in Arabic as partial and more conversational than academic.

As a Language and Meaning participant, Ms. Beydoun stood out as a teacher who was concerned with establishing a strong understanding of the Functional Grammar material for herself so that she could transmit the concepts accurately to her students. To that end, she and the other second grade/resource teacher, Ms. Shariff,⁸ reported meeting several times during evening and weekend hours to review lesson plans and other materials. This extra effort paid off, as Ms. Beydoun was observed to be one of the strongest implementers in the group of teacher-participants. Although the argument unit

⁸ Ms. Shariff began the school year as a second grade classroom teacher, but transitioned mid-year to a resource position. She was present in this capacity from time to time during the unit, working with a group of three newcomers who needed more Arabic-language support. Ms. Shariff's area of emphasis while studying for teacher certification was Language Arts and Bilingual Education, and her highest degree attained was a Master's in Bilingual Education. Her first language was Arabic and she reported being partially fluent, with some oral academic language. She had been teaching for nine years and taught first and second grade prior to becoming a resource teacher.

lessons were taught by members of the Language and Meaning research team, Ms. Beydoun was present throughout the unit and provided invaluable support in terms of making sure tasks and discussions were grade-level appropriate and assisting in managing classroom activities.

Like all teachers in the district, Ms. Beydoun's daily routine included extended literacy blocks in the morning. These literacy blocks were organized around the "Daily Five" program, in which students participated in one of five activities: Read to Self, Read to Others, Word Work, Work on Writing, and Listening to Reading. Students made their own choices regarding how to spend the Daily Five time, but were supposed to cycle through all of the options in the course of a week. When this 45-minute period started, the children decided what they would do for the day and spread out across the room. The time was self-directed, and there was not necessarily any shared content. Some students chose to sit at their desks, which were arranged in a long U formation facing a white board on one end of the room. Some sat at a small table or in a cushion-lined reading nook located at the other end of the room; others sat on the floor throughout the classroom. The few "newcomer" students who needed extra English language support either worked with Ms. Shariff or followed a language program on the computer, and throughout the literacy block Ms. Beydoun worked with one small reading group at a time in a corner of the room to the side of the white board. Our argument unit took place during this time of the day in place of the usual Daily Five activities (though most of the lessons required that the literacy block be extended somewhat).

An Overview of Lessons in the Argument Unit

The following section provides an overview of the instructional context in which the writing data were produced. The unit consisted of five lessons, which totaled approximately six and a half hours of instructional time over five days. This section describes the step-by-step unfolding of those lessons.

Day 1, Main Lesson (45 minutes; 15-minute follow-up). The primary goal for the first day of the unit was to give students initial exposure to the argument genre and its stages. Students would ultimately write arguments about whether or not grizzly bears, a threatened species, should be put in zoos. Because they would have a limited amount of evidence to draw on in discussing that issue, the materials for the introduction to the genre were built around a different, roughly parallel case, that of bald eagles, to provide an example that could serve as a model for the ways of thinking, talking, and writing that the focus on the grizzly bears would require.

The lesson began with a reading of a projected one-page text about bald eagles (see Appendix E), led by Ms. S. The text was designed by the research team to read as a general information text, but included several strategic points that could be used as evidence for a claim. The first half of the text included a number of facts about and characteristics of bald eagles. Most of these facts could be used to argue that zoos either were or were not appropriate environments for eagles. The second half of the text told the reader that bald eagles were once an endangered species and the reasons why this was so, as well as establishing that the bald eagle population was brought back to safer levels as a result of various human interventions, including putting them in zoos. Reading of the text

was supported by interaction through questions to check students' comprehension and review prior knowledge.⁹

After this interactive reading, Ms. S. introduced the problem to the students: putting bald eagles in zoos ended up helping save them, but at the time there was some controversy regarding whether or not it was a good idea. Using this problem as a starting point, students worked together to review their copies of the text and generate lists of evidence about why zoos were good for eagles and why they were bad. They then shared their findings with the class, which Ms. S. added to collective lists on poster paper at the front of the class. The concept "evidence" was introduced implicitly during this activity (i.e., used to describe the facts being added to the lists, without establishing it as a stage for the writing). Students were also asked to think about their opinions on the issue. Once the lists were complete, Ms. S. told the students that at the time of the controversy around putting bald eagles in the zoos, many children were writing about what they thought should happen. She then presented a model argument in support of putting bald eagles in zoos, purportedly written by one of these children. In fact, this model, along with the anti-zoo version that would be presented the next day (see Appendix F for both models), had also been written by the research team. Both texts included all of the elements of the argument genre being scaffolded and were written at roughly the same level of complexity at which the second graders were expected to write. Multiple models of the argument genre were generated in order to validate both sides of the argument and to show how evidence from the same source text could be marshaled in support of either claim.

⁹ Students had just finished a unit about the Alaskan River Ecosystem in which they learned about bears, eagles, salmon, and their habitat; this unit also featured a focus on the effects of human intervention on different ecosystems.

After reading through the pro-zoo model, Ms. S. displayed a poster listing the stages of an argument and a brief description of each. This official introduction of the genre incorporated the descriptions with a stage-by-stage deconstruction of the model. For example, in introducing the Description of Problem stage, students were asked to talk to a partner to decide where they thought the author was describing the bald eagle problem. Metalanguage used at this point included Description of Problem, Claim, Evidence, and Counterargument. The model's Reason, Evaluation of Counterargument, and Restatement of Claim stages were discussed without using the corresponding metalanguage; rather, Ms. S. framed her questions for the students in terms of the purpose of these stages (e.g., "Does the author agree that [the Counterargument presented] makes it a bad idea? Why?" for the Evaluation of Counterargument). Upon finishing the deconstruction of the model, students were told that they would be doing their own writing the next day, and the lesson concluded.

Day 1, Follow-up lesson. Later the same day, Ms. Beydoun decided to conduct a 15-minute review of the organization of the genre stage by stage. This review, like the last part of Ms. S.'s lesson, moved back and forth between reading descriptions of the stages from the poster and finding and highlighting relevant sentences in the model text, which was also projected for students to see. Metalanguage used during the follow-up included Claim, Evidence, Counterargument, and Evaluation. The "Reason" label was not used at all on the first day. Evidence and Reason were grouped together in discussion and highlighting, essentially treated as one unit. In both the main lesson and the follow-up, the word "argument" was used several times in verb form but only infrequently to describe the writing; both Ms. S. and Ms. Beydoun opted to use the label "Take a Stand"

to establish a connection with a persuasive text type that the students were already familiar with.

Day 2 (1 hour, 25 minutes). Ms. S. and I (Ms. K.) co-taught the second day of the argument unit, during which students began to consider the grizzly bear problem and the evidence they had available to support an argument. The day began with Ms. K. reviewing elements of the previous lesson. Students demonstrated strong recall of the bald eagle information, supplying facts and answering questions with minimal prompting. Their recall of the genre-specific information was weaker. They were not able to volunteer the stage labels on their own, but responded better when Ms. K. approached the material by asking them to define a particular stage or identify it in the pro-zoo model. After activating students' prior knowledge via the review session, Ms. K. presented the second, anti-zoo model argument. After an initial read-through, Ms. K. led a discussion in which students worked to identify the stages in this new model. It was in the context of this discussion that the Reason metalanguage was introduced. It had been established on Day 1 that Evidence came from the text, and this point was used to contrast Evidence with Reason, which involves the author's own ideas about why the chosen piece of evidence supports the claim being made. This "Evidence is from the text, Reasons are from your head" distinction, which would be drawn on many times throughout the unit, was also used to draw attention to the difference between the two sentences in the Counterargument/Evaluation of Counterargument paragraph.¹⁰ The Counterargument stage is, in essence, evidence that someone on the other side of the argument might use, while the Evaluation of Counterargument involves the author's unique response to why

¹⁰ The term "counterargument" was used at this point, but "Evaluation of Counterargument" was not introduced here either.

that evidence is not sufficiently convincing. Reiterating the evidence-text connection was important for the second graders, whose experience with related “Take a Stand” writing was opinion-based.

Following discussion of the anti-zoo model, Ms. S. took over and introduced the Grizzly Bear Fact Sheet (see Appendix G), a one-page list of facts about grizzly bears that would serve as the main informational text for the students’ arguments. The information on the Fact Sheet was similar to the text on bald eagles: the first five of seven bullet points presented characteristics of grizzly bears, and the last two established that the bears may become endangered and why. Ms. S. projected the Fact Sheet and read through it with the students, asking comprehension-checking questions and reviewing concepts that students were already familiar with. Ms. K. then led the students in an evidence identification activity and Ms. S. transitioned into a note-taking role. Ms. K. established the problem—whether or not grizzly bears should be put in zoos—and worked with the students to review the Grizzly Bear fact sheet point by point, considering whether each fact could provide evidence relevant to the issue and, if so, which claim it could help support. Students consulted with one another and then shared back with the group; once a consensus was reached, Ms. S. recorded the evidence on one of two posters (each representing a possible claim). During this conversation, Ms. K. and the students practiced generating reasons orally (prompted by questions such as “Why does it mean they shouldn’t they put grizzly bears in zoos if they need lots of space?”). We considered it important to keep the logical connections between claim, evidence, and reason at the forefront of the children’s minds; further practice with reasons would take place on Day 3.

Day 3 (1 hour, 10 minutes). The third lesson in the unit was also co-taught. The lesson objectives were to co-construct the Description of Problem stage, to review and practice finding/generating evidence and reasons, and to allow students to write their own Evidence and Reason. Ms. K. began the lesson with a review of what students had learned about grizzly bears. Students were highly engaged during this review, raising their hands to add to the discussion and frequently calling out answers, and their recall of the grizzly bear information was strong.

When the review had concluded, Ms. S. took over and worked with students to co-construct the Description of Problem. Ms. S. wrote on a projected piece of paper while students worked to select relevant facts about the grizzly bear situation and help with spelling. The final Description of Problem read “*Grizzly bears are becoming endangered. There used to be 50,000 grizzly bears but now there are only 1,000 left in the USA.*” Because the level of difficulty involved in summarizing and selecting the appropriate level of detail for an effective Description of Problem was considered too challenging for the second graders to take on in the context of this unit, the co-constructed text was shared amongst the students. Ms. S. told the students that they could copy the Description of Problem when they wrote their arguments, but they still had to “make a big decision.” Students had been considering their own opinions on the issue since the day before. After a brief refresher with Ms. K. on Claim, Evidence, and Reasons, students were asked to tell a partner what their claim was currently and why. Discussion on this topic was then opened to the whole class. Students from both sides of the issue were invited to contribute, as we wanted to legitimize both positions and allow the children to provide one another with different points to consider.

Once a number of students had shared their thinking, the class was divided into groups according to their current thinking regarding claim. Students in favor of putting grizzly bears in zoos worked with Ms. S, and those against doing so sat with Ms. K. The procedure for both groups was similar. First, the groups reread and discussed the evidence poster generated for their claim on Day Two, adding any pieces of evidence the students felt were missing. Then, taking one piece of evidence at a time, the students practiced explaining why the evidence meant that grizzly bears should or shouldn't be put in zoos, sharing first with a partner and then with the rest of the group. The teachers moderated the conversation and recorded the co-constructed reasons on a new poster. The only substantial difference between the two groups concerned practice with modality¹¹. In the prior Functional Grammar unit, students had worked on making predictions using words that told the reader “how sure” they were, including *probably, may, might, could, maybe*. The group that felt grizzly bears should not be put in zoos incorporated these words into their practice with Reasons, discussing during co-construction how sure they were about what they were saying and which of these words would be appropriate to use. The group that felt grizzly bears should be in zoos had this conversation without a focus on modality; rather, this group was told at the end of the activity that when writing their Reason, they should “think about whether this is something that will happen for sure, or maybe will happen, might happen.” Finally, students returned to their desks and were given approximately 15 minutes to begin writing their Evidence and Reason.

Day 4 (1 hour, 25 minutes). The fourth lesson began with an opportunity for students to finish writing their Evidence and Reason. To help them with their writing, this 25-minute portion of the lesson began with a review of the prompt—*Should people put*

¹¹ This difference was not planned.

grizzly bears in zoos? Use specific evidence and reasons to explain or why not—and a chance for students to reiterate (or change, if they so desired) the position they were taking. As students began to write, Ms. S., Ms. K., and Ms. Beydoun monitored progress and answered questions. In reviewing her students’ writing, Ms. Beydoun noticed that some were having difficulty formulating their Reasons. In consultation with the research team and in keeping with an established classroom practice of offering “sentence starters”, students’ writing time was paused for a five-minute discussion on beginning the Reason with the phrase “If we put grizzly bears in zoos...” This option was demonstrated with several student examples, and then students returned to writing. Later, Ms. S. led another brief discussion, this time reviewing “probability words”—the metalanguage used to describe the work on modality described above—and talking through two students’ examples to illustrate the concept.

When the initial writing portion had concluded, students were split into two groups according to their positions on the grizzly bear issue and the groups moved to opposite sides of the classroom for an interactive activity. This activity was designed to bring the challenging Counterargument/Evaluation of Counterargument stages to life, allowing students to embody its social purpose; it also served as an oral rehearsal of the ideas they would later commit to paper. The structure of the activity was as follows: A student representing one claim or “team” stood up and, with the help of a microphone, read his or her Claim, Evidence, and Reason to the class. The members of the other “team” then consulted with one another and responded with their Evaluation of the other side’s argument, beginning with the phrase “I disagree because...” This sequence was repeated beginning with the other claim.

Two issues arose in the course of the 20-minute activity. The first was that while the activity had been envisioned as a point/counterpoint exchange, the structure quickly became one of point for/point against. This was likely partially due to the cognitive challenge posed by what we were asking of the students. It also had to do with the second issue that emerged, namely, that there was very little variation in the ideas generated by the group that was against putting bears in the zoos. Upon later reflection it became clear that this was due to the information available in the source material. This issue would turn out to have a significant impact on the final written arguments, and is discussed in detail below. Overall, however, it was clear that the students quite enjoyed the activity, and interview data collected once the unit had concluded show that it enhanced several focal students' understanding of the Counterargument (see discussion of focal student interviews later in this chapter).

This activity was followed by modeling and co-construction of the written Counterargument. Ms. S. modeled two Counterarguments, writing on a projected piece of paper as she discussed what she was doing with students. The phrase "Some people say..." was reintroduced here to begin the Counterargument. Though it had been present in the model arguments read in previous lessons, it had not been used in the oral rehearsal because of the way that activity was structured. At the conclusion of this discussion, students took the last 15 minutes of the class to begin to write their Counterarguments.

Day 5 (1 hour, 25 minutes). The goal for the final lesson was to give the students time to combine the elements of their arguments together into one final draft. Ms. Beydoun took the lead role in the review that started the lesson. She first referred back to one of the bald eagles models to review the stages, reinforcing the connection between

that example and the poster at the front of the room listing the stages. She then used these resources to co-construct an argument with the students stage by stage, presenting the purpose (“what the reader wants to know”) and eliciting students’ examples to write the argument on a lined piece of paper that was projected for all students to see. Ms. Beydoun added a focus on paragraph breaks, which was not part of the FG unit but related to other classroom goals and practices.

After completing the co-constructed argument, there was one last brief review of the stages and how they fit together, and then the students went to their desks to write for the duration of the class period. As they wrote, they had support from Ms. S., Ms. K., and another member of the research team who had been present throughout the unit. Ms. Beydoun worked with a small group of students who needed additional support, in many cases because they had not finished writing the stages begun during previous lessons. Some students were able to combine the elements of their arguments with little to no guidance, while others had questions or needed to be reminded to check the order of their stages against the poster. Little attention was paid to the Restatement of Claim in the initial review, and in the end most students needed a reminder to include it, which they did by simply copying their initial claim at the end of what they had written.

Pre- and Post-Instruction Writing

This study’s design included pre- and post-instruction argumentative writing tasks intended to gauge students’ unscaffolded performance on a prompt similar to the one used in the argument unit before and after participating in the unit’s lessons. The prompt was the same for the pre- and post-instruction tasks, and the same prompt was used in second and fourth grades. The subject matter was drawn from the FG unit that students at

both grade levels had just completed. The scientific content for that unit focused on the impact of human activity on different ecosystems. During one of the lessons, students read and discussed a text about the different parts of the sandy beach habitat. One of those parts is *beach wrack*. Beach wrack is piles of seaweed, shells, trash, and dead animals. It provides a home for a number of organisms, and organisms from other parts of the sandy beach habitat depend on the organisms that live in the beach wrack for food. Before writing their arguments, students filled out a three-question, multiple choice quiz designed to refresh their memories of this content. The prompt students were then asked to respond to in the pre- and post-instruction tasks read:

In Florida, there are many hotels along the sandy beach. Every morning, the people who work at these hotels go out and remove the beach wrack so the hotel guests can swim on a clean beach. Should the hotel workers take away the beach wrack? Use specific evidence and reasons to explain why or why not.

It should be noted that students did not at any point read additional materials about hotels removing beach wrack from the shore. Rather, their responses were expected to be informed by prior exposure to information about the consequences of human interference in different ecosystems. The position that removing the beach wrack is undesirable because it would have a negative impact on a number of organisms would be easier to support scientifically, but if students were inclined to argue from the standpoint of more everyday reasoning they could (and did) make the case that removing beach wrack was a good idea for the hotels and their guests because a cleaner beach is more enjoyable.

The pre- and post-instruction prompts were administered by classroom teachers, who followed instructions provided by the research team. Arguments produced in

response were evaluated by independent raters who applied a primary trait rubric. The rubric, which had been used in Spring 2012 by all Oak Grove teachers, was an adaptation of the 6+1 Traits Writing Model (Culham, 2005). Under this model, writing is scored with rubrics emphasizing six key constructs: Ideas and Content, Organization, Voice, Word Choice, Sentence Fluency, and Conventions. To evaluate students’ pre- and post-instruction arguments, raters followed the procedure described in detail in Chapter 3 (scoring materials can be found in Appendix B).

Inter-rater agreement. To determine reliability for independent scoring, percentages of exact and adjacent agreement (defined as scores within one point of one another on the rubric’s six point scale) were calculated. Combined agreement ranged from 60% to 100% for the pre-instruction writing and from 68% to 96% for the post-instruction writing. Table 4.1 displays the combined percentages for exact and adjacent agreement; for a breakdown of exact versus adjacent agreement, see Appendix N.

	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
Pre	64%	83%	60%	92%	100%	96%
Post	68%	84%	76%	96%	96%	92%

Table 4.1. Inter-rater agreement (exact and adjacent) for second grade pre- and post-tests

Agreement was generally lower for writing done before students’ participation in the argument unit compared to writing done after, likely due to higher variation in arguments produced before being exposed to FG instruction and building a shared idea of the features of this genre. As described in Chapter 3, arguments for which scores on any trait differed by more than one point were referred to a third rater. A review of the third set of scores for cases where exact or adjacent agreement fell below 30% for either pre- or post-instruction writing—in this case, Ideas & Content and Voice—found that in the overwhelming majority of cases, the third rater’s scores were within one point of those of

one of the two original raters. Of the 12 pre-tests and 10 post-tests referred to a third rater, there was just one instance in which there was a discrepancy of more than one point for the Voice trait between the third score and the initial two. The “tie-breaker” round of scoring, then, tended to reinforce the scores produced in the initial round of scoring and functioned to provide additional data to round out the final scores used in analysis, which were an average for each trait of all the ratings assigned to a particular piece of writing. The next section presents a discussion of trends in these final scores.

Results of scoring. In the second grade, 20 students, or 80% of the class, improved on some or all of traits measured by the independent raters’ rubric. The five students whose performance decreased represented a range of language proficiency and academic performance subgroups. Two students in this group were in the At or Above Grade Level reading group, one was in the Some Risk group, and two were labeled At Risk. Similarly, two students tested at ELPA Levels 1-2, and the remaining three were at Levels 3, 4, or 5.

Of the students whose performance improved, 11 were given higher scores on all traits. The remaining nine improved on some traits but their scores decreased for one or more others. Seven of these students saw a decrease in their Conventions score; for three of them, this was the only trait for which they received lower scores. For the remaining five traits there were only one or two instances of lower scores each.

There did not appear to be either “ceiling” or “floor” effects in the second grade pre-/post-test data. The students who performed best on the pre-instruction task (those with who generally had scores of 3.5-4 across traits) were essentially equally likely to receive higher or lower scores; while some of these initially high performers did see a

decline in performance on the post-instruction task, others with similar initial scores were rated even higher on the second task. The lowest-performing students had pre-test scores in the 1-1.5 range across traits in the second grade. Though in a very few cases it would not have been possible for scores to decline, this was not an issue: children who initially scored among the lowest in their classes uniformly showed improvement.

In several cases, it was also the students in the lowest language and reading groups whose writing improved most in terms of the traits that aligned most strongly with the goals of the instruction. Specifically, the FG program emphasizes ideas and how language is used to convey and organize them, positing that while conventions are not without value, attention can be paid to them once the ideas around which a piece is built are in place. Table 4.2 shows the average change in points for each trait in the second grade, organized by subgroups. On the Ideas & Content trait, students at ELPA Level 5 and those in the At Risk reading group made average gains of over half a point, more than any other subgroup. The Level 5 language subgroup also showed the strongest gains for Organization, though the At Risk group made gains close to the class average. Conventions scores for these lower groups generally decreased. Second graders in the At/Above Grade Level reading group, on the other hand, show average gains across all six traits. This suggests that while increased focus on the FG content may draw lower-performing students' attention away from surface-level features, the structure presented in the argument unit may in fact have streamlined the writing process for others, allowing them to pay more attention to these "finishing touches". (For additional discussion of these results, see Chapter 6; for individual students' scores, see Appendix O.)

	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
<i>Overall</i>	0.41	0.39	0.26	0.23	0.22	-0.02
<i>ELPA</i>						
1 & 2s (n=5)	0.43	0.40	-0.03	0.07	0.12	0.16
3s (n=6)	-0.14	0.14	0.28	0.47	0.14	-0.03
4s (n=5)	0.40	0.40	0.23	0.33	0.27	0.00
5s (n=6)	0.56	0.56	0.36	-0.03	0.28	-0.11
<i>Reading</i>						
At/Above Grade Level (n=10)	0.47	0.42	0.28	0.33	0.35	0.22
Some Risk (n=7)	0.11	0.39	0.00	0.11	0.30	-0.17
At Risk (n=10)	0.56	0.37	0.41	0.20	0.04	-0.18

Table 4.2. Average pre- to post-test changes in scores by subgroup, 2nd grade

While this study’s design limits the claims that can be made regarding the effectiveness of the instruction itself, the primary trait scores provided by independent raters demonstrate that using a form of assessment that is typical in their context, Oak Grove teachers rated most student writing produced after participation in the argument unit higher on most traits than the writing produced before the instruction, with the strongest gains typically observed for the dimensions emphasized by the FG approach. The following section reports on my own evaluation of student writing using an SFL approach to analysis.

Written Arguments

After the unit had concluded, students’ arguments (n=25) were transcribed and standardized to prepare them for use in linguistic analysis. Several analyses were performed in order to characterize the clause- and stage-level features of students’ argumentative writing. The results were further analyzed to produce descriptive statistics for the performance of the class as a whole as well as for four language proficiency level

bands and three reading level groups.¹² In the following discussion, examples from student writing will identify the sample number as well as the ELPA and reading levels of the writer. ELPA levels include 1-2, 3, 4, and 5; the three students for whom no spring scores were made available are designated as NSS. The reading levels were At or Above Grade Level (AAGL), Some Risk (SR), and At Risk (AR). These were the reading levels used by Ms. Beydoun, and she indicated that they generally corresponded to the report card categories of Meeting Expectations, Progressing, and Concern.

Trends: Clause-level analysis. Arguments were first parsed into component clauses and T-units, and any embedded clauses were identified. A T-unit (or minimal terminable unit) is a main clause and any attached subordinate clauses (Hunt, 1965). The T-unit, which provides an approximation of syntactic complexity, is a widely-used measure in linguists' studies of language and writing development (Schleppegrell, 2008; Wolfe-Quintero, Inagaki, & Kim, 1998); including counts for T-units and individual clauses allows for this work to be in dialogue with SFL-oriented research as well as work grounded in other perspectives (e.g., Loban, 1976; Perera, 1984; Myhill, 2007, 2008). Using T-units rather than sentences as a unit of analysis is especially advantageous in working with young children's writing, as T-units provide for systematic analysis even if, as occurred many times in this sample, writers do not consistently use punctuation to mark clause or sentence boundaries.

The relationship between clauses and T-units can take a number of forms. For example, the T-unit was conceived of as the minimal unit that could be considered a

¹² The only student scoring at ELPA Level 1 (Advanced Proficient) was grouped with the four students scoring at Level 2 (Proficient) to facilitate statistical description; Levels 3-5 (High Intermediate, Low Intermediate, and Basic; n=6, 5, and 6, respectively) comprised the remaining three bands. The reading level groups were those used in the classroom: At/Above Grade Level (n=10), Some Risk (n=6), and At Risk (n=9).

complete sentence. Many sentences can be analyzed as a single T-unit and a single independent clause, with no embeddings:

They should put grizzly bears in zoos.

This, of course, is not always the case. Consider the following sentence, which consists of two separate clauses (boundaries marked with slashes), which together form a single T-unit:

Put them in zoos / because no humans can kill them.

The next sentence also represents one T-unit, but includes three embedded clauses (set off with brackets, boundaries marked with slashes):

Grizzly bears need a lot of space [to live / and hunt / and play].

The mean number of total clauses for the second-grade arguments was 10.52, with a range of 7 to 13 clauses and a standard deviation (SD) of 1.39. The total number of T-units ranged from 4 to 9 with an average of 6.12, SD 0.97. Clauses per T-unit ranged from 1.33 to 2.20, SD 0.29; mean T-unit length was 1.74 clauses. Most students used no embedded clauses, though a few used as many as three; the mean number of total embeddings per argument was 0.80 (SD 1.04), with an average of 0.14 embedded clauses per T-unit (ranging from 0 to 0.50, SD 0.17).

Only minor differences were observed between students at different language proficiency levels. For example, students testing at ELPA Level 3 used an average of 1.67 embedded clauses, more than twice the rate of any other language proficiency group, but this figure does not represent any identifiable trend. Furthermore, mean numbers and length of clauses and T-units were roughly the same for all groups: the difference

between the lowest- and highest-scoring groups on these measures was only about three tenths of a point. For the group under study, language proficiency did not appear to be related to performance on these measures. Nor, for its part, was reading level. Students classified as At/Above Grade Level, Some Risk, and At Risk performed roughly the same on these measures. It is possible that some differences would emerge given a larger sample, especially as regards embedded clauses. It is likely, however, that the structured nature of this stage-based approach—especially with younger children writing in a new genre—functions to limit the diversity that might be evident in responses to a less structured task. Older children (such as the fourth graders whose writing is discussed in Chapter 5) and those with prior experience with the genre would be expected to demonstrate more variety in their responses. Another possibility is that the level of support provided was enough to bolster the performance of all students, including those who under other circumstances may have produced simpler or shorter texts. The following section maintains a clause-level focus, but shifts to analysis of how genre-specific meanings are construed by students' language choices.

Trends: SFL metafunctions. SFL models of language posit that in every instance of language use, three kinds of meaning are realized simultaneously through the lexicogrammatical choices made by the writer or speaker. These types of meaning, referred to as metafunctions, consist of the ideational, the interpersonal, and the textual (Halliday, 1994). Ideational meaning has to do with the presentation of ideas. Ideational resources including participants, processes, and circumstances that tell us what is being talked about. Interpersonal meaning concerns the social relationships between speakers/writers and their listeners/readers. In written language, grammatical mood and

modality are used to convey interpersonal meaning. Textual meaning is about the organization of information. Linguistic resources used in the structuring of a text include those to do with cohesion and coordination, as well as choices regarding *grammatical theme* (the first linguistic element in a clause, representing a “point of departure” for the clause or message).

Language choices associated with the three metafunctions vary according to purpose and context. Genre is, in essence, the way these choices are organized by a given culture into “recurrent configurations of meaning” (Martin, 2009, p. 12). While the “rules” for performing genres are not iron-clad, every genre invokes a particular constellation of contextually appropriate meanings, and a basic description of the way resources are typically coordinated can be a useful starting point for learners. It can also serve as a tool for writing analysis, revealing whether what students have written is consistent with the goals of the genre. In the present study, such an analysis showed that this was the case for the second graders as a whole.

Ideational resources. The argument genre presented in the unit is characterized by the use of generalized participants (i.e., “grizzly bears” as a class rather than naming a specific participant), a pattern reflected in student writing. Participant analysis showed that all reference chains used by the students concerned generalized participants (grizzly bears, cubs, zoos, zookeepers, people, cages, habitat, space). The students’ writing also conformed to expectations that the majority of processes used would be *being* and *doing* processes, used in this case to describe grizzly bears, report on what they do, and offer opinions on which actions should or should not be taken and why. Students used an average of 7.18 *doing* processes (0.68 per clause, or PC) and 2.10 *being* processes (0.20

PC) in their arguments, compared to 1.09 (0.10 PC) and 0.59 (0.06 PC) *saying* and *sensing* processes, respectively. The *saying* process used by all students was an artifact of the way the genre was scaffolded, with the phrase “Some people say...” used to introduce the Counterargument. The small minority of students who used *sensing* processes employed them to describe bears’ needs and preferences (e.g., “Grizzly bears *don’t like* to be near humans”).

Some movement toward a higher concentration of *being vs. doing* processes was observed as language proficiency level increased. *Being* and *doing* processes are the most commonly-used process types in texts about scientific topics because their meanings concern phenomena in the world rather than what people think or feel. As writing on these topics becomes more advanced, there tends to be a higher concentration of *being* processes in particular, as these are used to make generalizations, discuss relationships between ideas, and so on, in addition to being related to grammatical metaphor¹³ (Christie & Derewianka, 2008). The highest-proficiency group used an average of 6.4 *doing* processes (0.60 PC) and 2.8 *being* processes (0.26 PC), while the lowest-proficiency group used an average of 8.0 *doing* processes (0.77 PC) and 1.5 *being* processes (0.15 PC). Close inspection of the way the processes were used suggests that these differences are not significant. *Being* processes are used to make predictions about bears being safe or having enough space, but these ideas are also expressed with *doing* processes: “If we

¹³ Halliday (e.g., Halliday & Matthiessen, 1999) describes the development of grammatical metaphor as the movement from congruent to incongruent realizations of experience; in other words, a shift from representing experience in the grammatical forms which evolved to express these meanings—in which, for example, actions are realized as verbs (“People shoot grizzly bears in the wild”)—to a more metaphorical one, which involves a repackaging of information such that an action can be realized through a nominal group: “The shooting of grizzly bears in the wild is a threat to their survival.” Because this compacted information is the basis for abstract, technical language, Christie and Derewianka hold that control over grammatical metaphor is “fundamental to the very nature of educational processes in the higher levels of schooling—the construal of experience into specialized domains and the reasoning about experience in abstract, logically developed terms” (2008, p. 25).

put grizzly bears in zoos, they will be safe” vs. “If we put grizzly bears in zoos nobody will hurt them.” use *being* and *doing* processes, respectively, to communicate ideas that are very close in meaning. Similarly, the same Claim could be phrased with either process type with only a slight change in meaning: “Grizzly bears should not be in zoos” vs. “We should not put grizzly bears in zoos.” The only difference in meaning regards human agency, and there is no evidence in either written or observational data that this was a conscious choice on the part of the students.

The ideational resources used to establish logical relationships in the student texts were also consistent with the goals of the genre. Conjunctive resources used spontaneously were overwhelmingly those of *cause* or *reason*, usually the connector “because”. There were 48 instances of this type of usage in the data set. Students were never explicitly told to use this connector, but it is crucial in building ideational meaning in a well-developed argument: “because” signals an explanation of the logic underpinning the idea it follows. Many students also included the connector of *condition* “if” in making predictions to support their claims, as modeled in the unit (“If we put grizzly bears in zoos...”). There were 21 such instances in the data. Other connectors, used much less frequently, were those of *contrast* (“but”, used twice) or *addition* (“and”, used four times), used to coordinate clauses¹⁴.

A minor pattern was observed in terms of increased connector usage as language proficiency level increased. Students testing at ELPA Levels 4 and 5, the two lowest levels, used an average of 0.27 conjunctive meanings per clause. This average grew to 0.30 PC for students at Level 3 and 0.32 PC for those at Levels 1 and 2. In addition, as

¹⁴ Additional instances of “and” being used to link phrases were noted in the sample but not included here, as the focus on connectors has to do with clause, rather than phrase, linking.

reading level increased there were slight increases in overall use of conjunctive meanings as well as in the use of connectors of *cause/reason*. Students labeled At Risk used 0.25 conjunctive meanings per clause on average and those labeled Some Risk used 0.29 per clause. Students reading at or above grade level used 0.31 conjunctive meanings per clause. For the same reading groups, the mean number of connectors of *cause/reason* per clause were 0.16, 0.19, and 0.20, respectively. While sample size precludes drawing strong conclusions from these results, there is some indication that growing familiarity with written English brings with it an increased sensitivity to the kinds of meanings that are valued in a particular context.

Interpersonal resources. Analysis of interpersonal resources involved examining *grammatical mood* and use of *modality*.

Mood. The second-grade writing data included declarative, interrogative, and imperative clauses. Only the first would be considered typical of the argument genre, and indeed, almost all of the clauses were of this type. Just two students departed from this pattern, one using several interrogatives and the other using an imperative:

...[C]an a zoo be 20 [city name]s? No! (Sample 86, 3/AAGL)

Put bears in zoos... (Sample 2, 4/AAGL)

The writer who used the interrogative tested at ELPA Level 3, and the writer who included an imperative tested at Level 4. Both were reading at or above grade level. These departures from the expected *mood* type worked in different ways to strengthen the writers' messages. The author of Sample 2 drops the modal *should* used elsewhere in the argument, removing a layer of rhetorical buffering from the recommendation. Sample 86's writer engages the reader directly with a rhetorical question, repeated several times

throughout the essay. In Sample 86, several versions of the interrogative clause presented above appear. In the Evaluation of Counterargument the question is used to explain why building bigger cages would be an inadequate solution to the grizzly bears' problem. At that point it goes unanswered, thereby engaging the reader directly by asking her to supply one. In the Restatement of Claim, the answer is supplied and emphasized with an exclamation point, leaving no doubt about the writer's position or the strength of his feelings. While these variations from the simple declarative presentation of the genre are atypical, neither students' *mood* choices are inappropriate, and may in fact indicate a more sophisticated understanding of the range of choices available to them in composing an argument.

It is also interesting to consider the relatively low incidence of interrogative and imperative clauses in light of students' exposure to these phenomena in other reading and writing contexts. Writing in elementary school frequently emphasizes the concept of "voice", typically defined in this context as making writing interesting, grabbing readers' attention, showing "personality", and the like. In addition, although scientific writing typically values an authoritative, detached voice, scientific texts in the lower grades frequently make use of a variety of mood choices in an attempt to engage their young readers. That students largely chose to use declarative clauses, then, indicates that they were sensitive to the mood type used in the models and co-constructed arguments even though this construct was never explicitly discussed.

Modality. Three types of modality were presented in students' writing: modality of *obligation*, *ability*, and *likelihood*. All of the students used modality of *obligation*, the most strongly scaffolded of the three. This almost invariably took the form of *should*

and/or *should not*, generally appearing in the Claim and Restatement of Claim, and to a lesser extent in the Counterargument.

Slightly more than half (51.85%) of the second-grade writers used modality of *ability*, employing some form of the verb *can* in one of several ways. The most common purpose was to present what zoos or people can or can't do to remedy problems:

I disagree because the zoo *can* build a BIG cage. (Sample 71, 3/AAGL)

They *can't not* make bigger cages with 20 Oak Grove big. (Sample 45, NSS/AR)

Some instances were grammatically infelicitous and technically unnecessary, although elsewhere in their arguments the same writers also demonstrated the ability to use this type of modality appropriately:

...in the wild they *can* kill bears more... (Sample 106, 4/SR)

Finally, just one student used another form of *can* to end her essay:

But I say we *could* put bears in zoos. (Sample 128, 1-2/SR)

As Samples 45 and 106 make clear, many second graders are still learning how to deploy this linguistic tool appropriately. Even the variation in Sample 128, when looked at in the context of the entire essay, may not have been an intentional choice (it may have been an inadvertent substitution for *should*). Modality of *ability* was the only type for which there was a discernible trend in the data by ELPA level. 80.0% of students at the two highest language proficiency levels used modality of *ability*. Rates of usage at Levels 3, 4, and 5 were 66.67%, 40.0%, and 33.33%, respectively. In terms of reading level, students in the highest reading group were the most likely to use modality of *ability* (70.0%); students in the lower reading levels were 20-41% less likely to do so. This type

of modality is not a central feature of arguments and was not a focus of instruction. If it were, these data suggest a need for explicit support in this area.

Finally, 81.48% of the second-grade writers used modality of *likelihood*. Most students who did so used *will* or *won't*, generally in the service of making predictions about what would happen if grizzly bears were put in zoos. Although there had been some discussion during the unit of words that showed “how sure” students were about their predictions, there was very little variation from these highly polarized terms. One of the only exceptions built up modality using by modifying *will* or *won't*:

If we put grizzly bears in zoos *maybe will not* have a big cages. (Sample 153, 5/AR)

Taken together, these data show that the second-grade writers are beginning to build linguistic skills in the area of modality, though most still have a limited repertoire. It is possible that increased instructional attention to this point would produce more variety in students' writing. However, the writing analyzed here is by no means an example of poor performance: Modality is a feature that is known to develop later in writing than in speech and one that continues to pose a challenge for many second-language speakers of English even into their university years (Christie, 2012; Ventola, 1997).

Textual resources. As noted above, the *theme* is the first linguistic element of the clause. When the *theme* is simply the topic of the clause, it is referred to as *unmarked*:

Grizzly bears should be put in zoos.

A *marked theme*, in contrast, foregrounds something other than the topic of the clause (Halliday & Matthiessen, 2004). For example, writers may choose to highlight additional information by putting a dependent clause or prepositional phrase in Theme

position. Of all the clauses in the second-grade data, only two instances of marked Themes were found, both involving a circumstantial element in theme position:

In the zoo no one can kill bears. (Sample 2, 4/AAGL)

In the zoo cages are small. (Sample 18, 3/SR)

The preponderance of simple *unmarked themes* has less to do with choices related to the expectations of the genre—which have little to say about *theme*—than with the developmental stage of the writers. This point is discussed in greater detail below.

Summary. The linguistic choices made by these second-grade writers were largely consistent with both the demands of the argument genre and the kind of writing that is typical for writers of this age. Ideational meaning was built through reference to generalized participants; the use of mostly *being* and *doing* processes, in keeping with the need to present factual information in support of an argument; and through connectors of *cause/reason* and *condition*, which function to present explanations of argumentative logic and make predictions. Interpersonal resources related to grammatical *mood* and modality were also genre-appropriate: almost all clauses were declarative, and students used modality of *obligation* and *likelihood* to make recommendations and predictions. Other features of student writing were not unexpected, given the age of the writers.

Christie and Derewianka (2008) conducted a large-scale developmental study of children's writing from early childhood to late adolescence. Their findings, presented in terms of SFL metafunctions as well as traditional linguistic measures, provide a context for understanding the data in the present study. For example, the observation that use of marked themes is characteristic of more advanced writers and uncommon in the writing of students of this age helps establish the second graders' use of mostly *unmarked themes*

as developmentally appropriate. Other developmental findings common to that study and the present one include the high incidence of single-clause sentences or those that combine clauses of equal status, with occasional use of non-finite clauses of purpose, and a somewhat limited range of conjunctive resources.

In the next section, the focus of analysis turns to how students constructed meaning at the *stage* level.

Stage Analysis

Genres are described in terms of their social purposes, stages, and language features typically used to build meaning. The preceding section showed that the second-grade writing included developmentally-appropriate language features that were characteristic of the argument genre. To deepen our understanding of how these young students begin to construct the genre, however, it is necessary to place these observations in context. In this section, discussion will shift from an examination of linguistic features in isolation to a stage-by-stage consideration of the ideas they were used to construct. Stages in the SFL paradigm are phases of meaning, each with their own purpose and together building toward the overall goal of the genre. Martin (2009) explains stages as follows:

...[W]e cannot achieve our social purposes all at once, but have to move in steps, assembling meaning as we go, so that by the end of a text... we have ended up more or less where we wanted to be. (p. 12)

In designing the argument unit, we drew on other SFL descriptions of persuasive texts (Christie and Derewianka, 2008; Derewianka, 1991; Martin, 1989) and Toulmin's (1958) canonical description of argument and ultimately put forth a description of the

genre that included seven stages: Description of Problem, Claim, Evidence, Reason, Counterargument, Evaluation of Counterargument (hereafter called Evaluation), and Restatement of Claim. At the second-grade level, the writing of three of those stages was scaffolded to such a degree as to not warrant detailed discussion of the data. The Description of Problem stage was co-constructed and shared by all students. Likewise, the Claim and Restatement of Claim varied little either within or between arguments, with the obvious exception of overall position: 10 students wrote in support of putting grizzly bears in zoos, and 15 argued against it. The most meaningful data, then, come from the Evidence, Reason, Counterargument, and Evaluation stages.

The analysis below considers these stages in pairs: Evidence/Reason and Counterargument/Evaluation. In part this reflects the way the stages were treated in instruction. Reviewing classroom video for use of metalanguage shows that Evidence and Reason were nearly always discussed in conjunction with one another, while the concepts of Counterargument and Evaluation were fused to such an extent that the latter stage label was frequently not even used. More important—and likely related—is the fact that with such short arguments, there is very little time to build up significant meaning in a single stage (most of which consist of a single sentence). Thus, the social and rhetorical success of Evidence and Reason in these data are very much dependent on one another, and adhering to the single stage as the unit of analysis would give an inaccurate picture of students' performance.

Procedure. Transcribed arguments were first divided by stages. In the second-grade texts this was a straightforward process as stage boundaries were quite clear. All instances of stage pairs were grouped and placed into separate documents (e.g., one

document for Description of Problem and Claim, another for Evidence and Reason, etc.) These documents were then reviewed and stage pairs were labeled as Successful, Partially Successful, or Unsuccessful based on an evaluation of the extent to which the writer met expectations for each stage. The criteria used to make these assessments will be presented below as each stage is discussed. The designations reflected my evaluation of performance in absolute terms but also relative to the rest of the writing samples, and the labels should be understood in terms of what they reveal about additional support students may need. A “Partially Successful” label, in other words, may be applied to a stage pairing that, outside of the context of this analysis, would be considered an example of good second-grade writing (for more discussion on this point, see Chapter 6). To test for reliability, a member of the research team used the coding manual (see Appendix M) to assess 20% (n=5) of the second grade arguments. This process consisted of assigning Successful/Partially Successful/Unsuccessful labels to the Evidence/Reason and Counterargument/Evaluation stage pairs in transcribed arguments from which all identifying information had been removed. The exact agreement rate between my scores and the second rater’s was 80% for the second grade texts, and as such coding for this analysis was deemed reliable. The two instances of disagreement between raters involved minor differences in our perceptions of the writing, resulting in adjacent ratings for Evidence/Reason pairs.

Once these initial determinations had been made, stage pairs at each performance level were further reviewed in search of features they shared—to determine, for example, what made the Successful writing a success. This resulted in the identification of five

dimensions along which arguments across (and within) performance levels varied, to be discussed in detail below.

Evidence and Reason. 24 students included Evidence and Reason in their arguments. Of those 24, ten wrote pairings that were classified as Successful, 11 were Partially Successful, and just three students composed Evidence/Reason pairings classified as Unsuccessful. Table 4.3 presents the criteria used to evaluate Evidence and Reason in the second grade.

<i>Evidence and Reason</i>	
Successful	Presents a fact from the Grizzly Bears text and a logical explanation of how that information supports the chosen claim.
Partially Successful	A connection between evidence and claim is present but incomplete/insufficient, leaving more work to the reader than is ideal (“incomplete bridge”).
Unsuccessful	Lacks reference to Claim or anything that could otherwise be interpreted as a Reason, or adds commentary that does not make sense.

Table 4.3. Criteria used to evaluate Evidence and Reason

The ten students whose writing was Successful at this stage selected a fact from the Grizzly Bears text (Evidence) and gave a logical explanation of how that information supported their claim (Reason). In doing so, they successfully achieved the goals for these phases of meaning.

Three of the ten students included language that made their reasoning explicit, although this had not been explicitly modeled:

People are shooting the grizzly bears in the wild. They will be safer in zoos *because* people can’t kill them in zoos. (Sample 10, 1-2/AAGL)

People are shooting the grizzly bears in the wild. If they will put grizzly bears they will be safer in the zoos *because* they will not be killing the grizzly bears in the zoos. (Sample

91, 1-2/AAGL)

The remaining seven students elaborated less on their reasoning:

People kill grizzly bears. If we put grizzly bears in zoos they will be safe.

(Sample 140, 5/AR)

Grizzly bears need lots of space. If we put grizzly bears in zoos they won't have space to move around. (Sample 45, NSS/AR)

These students were also able to reason successfully about the consequences of putting bears in zoos without the need for additional explanation, and indeed, some may interpret extended explanation as redundant. The explicit “because” clause does not necessarily make for a better response, but serves as a mechanism by which students can do a self-check on their thinking, suggesting a potential pedagogical strategy (to be discussed in more detail below) for strengthening responses at this stage. Writers in this group were able to successfully execute these stages whether they let their reasons—all of which could be considered commonsense—stand on their own or included an extra move to explain their thinking.

Other students, however, struggled to make their logic clear and may in fact have benefitted from a push toward explicitness. 11 students wrote Evidence/Reason pairings that were only Partially Successful. Eight of these students provided what I call an incomplete bridge between their Evidence and Reason, one of two main areas of weakness in the Partially Successful and/or Unsuccessful examples. In these cases, there is enough of a connection between Evidence and Reason to be able to follow the argument, but more of this work falls to the reader than is ideal. For example:

Baby bears called cubs stay with their mother for 2 or 3 years. If we [put]¹⁵ bears in zoos the cubs will not know how to hunt. (Sample 63, NSS/AAGL)

¹⁵ Verb absent in original.

The argument here turns on multiple pieces of information that the writer failed to include, namely that mother bears teach their cubs to hunt in the early years of life and, ostensibly, that the writer believes that mother and cub would either be separated in the zoo or would not be provided living conditions suitable to hunting or learning to hunt. (Additional questions related to the author's understanding of hunting as an essential part of bear-hood could be asked, but for the purposes of this analysis I take at face value the belief that not learning to hunt is bad for cubs.)

Another, weaker example of the incomplete bridge:

Grizzly bears need lots of space. In the zoo the cages are small. (Sample 18, 3/SR)

Again, the reader can fill in the blanks without much difficulty, but in writing a successful argument it is incumbent upon the writer to clearly explain how the chosen evidence supports the claim. Missing here is an additional clause or sentence linking these two statements by noting how small the cages are—so small, ostensibly, as to be inadequate as homes for bears—or at minimum the modification of this clause to reflect the same idea (e.g., “too small”). Alternately, the Reason could have followed the “If we put grizzly bears in zoos, [predicted result]” format (see below).

This incomplete bridging is consistent with findings regarding elementary students' tendency to not repeat information they consider shared knowledge in oral arguments (Anderson, Chinn, Chang, Waggoner, & Yi, 1997). If students of this age believe that their writing is to be shared with their teacher and fellow students, it would be sensible on their part to assume that they and their readers have been exposed to the same information. Second graders are typically not yet skilled in writing for a distant audience; nor would they be expected to understand that a certain degree of explicitness

is valued in scientific writing. Add to this that the Counterargument “teams” activity strongly reinforced the notion that fellow students were the audience for the arguments, and it is unsurprising that some students included what Toulmin (1958) called implicit warrants. While it may be possible for a reader to follow an argument with implicit warrants, it seems wise to prepare students to identify their own warrants/reasons and present them explicitly. As Toulmin made clear in his treatise on argument genres, in the absence of warrants “it will become impossible...to subject arguments to rational assessment” (Toulmin, 1958, p. 100). Additional guidance toward explicitness at this critical juncture would almost certainly have strengthened the argument of the example above and others like it. In some cases the bridging problem was extreme—the three Evidence and Reason pairs that were classified as Unsuccessful had as their unifying feature the failure to establish any logical relationship between Evidence and Reason¹⁶—and in such cases additional guidance is almost certainly called for. Modeling and encouraging the use of the connector *because*, used as it was in the Successful examples above, could address this issue by serving as scaffolding in the traditional sense: as a tool put in place to help move students toward competence, and removed once they have more exposure to the learning target (Vygotsky, 1978).

Incomplete bridging largely concerns the expression of logic. Other problems were more directly related to language. For a number of students, linguistic infelicity was a confounding factor in the success of the Evidence and Reason stages:

They should not put them in zoos because they need lots of space. If we put bears in zoos

zoos never have a lot of space. (Sample 74, 5/AR)

¹⁶ A response characteristic of Unsuccessful Evidence/Reason pairings: Grizzly bears need lots of space. If grizzly bears are in zoos the baby bears won’t learn how to catch fish when they grow up. (78)

Grizzly bears need lots of space. If we put grizzly bears in zoos, their cages are too small.

(Sample 112, NSS/AR)

Both Samples 74 and 112 have clearly-linked Evidence and Reason stages, but are at the same time examples of the second common area of weakness, a problem found across language and academic proficiency levels: an improperly phrased or otherwise unclear conditional clause.

Conditional sentences contain two elements: a main clause that presents an action or situation, and a subordinate “if-clause” that presents the condition upon which the former is predicated. The kind of conditional sentence modeled for students involved the combination of *If* + a present tense verb (“If eagles are in zoos...”) in the subordinate clause and the use of *will* or another modal verb to make a prediction in the main clause (“...they will have to be around humans all the time.”) It appears that either the form of the conditional clause—known to be a challenging point grammatically—and/or the presentation thereof, which occurred late in the unit, contributed to a lack of clarity in some students’ Reasons. Although the prior FG unit had included practice with making predictions following an if-clause, students had not been asked to write their own original sentences combining these elements¹⁷. In the argument unit, examples of the if-clause “If we put bears in zoos” phrasing were present in models constructed and discussed throughout the unit. However, the notion that students could or should adopt this wording was only explicitly addressed once students had already begun (and possibly completed) writing their Evidence and Reason, when Ms. Beydoun suggested using it as a sentence

¹⁷ In this unit, students read about the Alaskan River Ecosystem and how different organisms within that ecosystem depend on one another. They completed the sentence “If there aren’t enough salmon...” by making a prediction about the effect that a reduction in the salmon population would have on other organisms, focusing on using modality to express how likely they felt the stated outcome to be.

starter mid-lesson on Day 4. Students who took this suggestion were effectively left to their own devices to construct an appropriately-linked main clause for the resulting conditional sentences. While a number of students were able to do so, others would likely have benefitted from more support.

In considering how to better support students to successfully produce this pair of stages, then, two potential areas of focus emerge, associated with two issues of logic and language that were common in these data. Encouraging the use of the connector of cause/reason *because* to make reasoning explicit would help address the tendency for some students to provide incomplete bridges between their Evidence and Reason. In addition, in future iterations it may be advisable to add to the unit a mini-lesson or some other opportunity to discuss and practice the linguistic structure of conditional sentences, a grammatical point that troubled several students.

Counterargument/Evaluation. The Counterargument/Evaluation stage pairs were again initially classified as Successful, Partially Successful, and Unsuccessful. Performance on this set of stages was not determined by how well students had achieved the goals of the Evidence/Reason pair. For example, 11 students produced successful Counterargument/Evaluation pairings. Six of these had also produced Successful Evidence/Reason pairings, and five had been Partially Successful. Similarly, the nine students who were rated as Partially Successful in writing this stage pair were evenly split between the three performance levels for Evidence/Reason. Finally, five students' attempts at this stage were marked as Unsuccessful. One of these students did not include the Evidence/Reason stage in his/her argument, and the remaining four were divided equally between those who had produced Successful and Partially Successful

Evidence/Reason pairings. Table 4.4 presents the criteria used to evaluate the Counterargument and Evaluation stages.

<i>Counterargument and Evaluation</i>	
Successful	Provides a piece of information that someone presenting another side of the argument might use for evidence and explains why the Counterargument does not change the writer's position.
Partially Successful	Presents Counterargument and attempts an Evaluation but does not adequately explain why the Counterargument does not change the writer's position (e.g., explanation is incomplete/logically weak).
Unsuccessful	Presents Counterargument but either does not include anything resembling an Evaluation or Evaluation is unrelated to Counterargument.

Table 4.4. Criteria used to evaluate Counterargument and Evaluation

The performance levels for the Counterargument/Evaluation pairings largely turned on the strength of the Evaluation. Constructing the Counterargument is a relatively simple task for students whose classroom experiences have been well-scaffolded, as was the case for these writers. They had been exposed multiple times and in multiple modes to the reasoning of students with claims opposing their own. Writing this stage was essentially a matter of reporting one of a limited range of potential Counterarguments. Evaluations, on the other hand, called for a much more individualized type of response and required more work on the writer's part to contextualize this stage in terms of both the Counterargument specifically and their argument more generally, explaining why the former did not sway their thinking. Students who were unable to establish a connection between their Counterargument and Evaluation were those whose responses were marked as Unsuccessful. Partially successful pairings were those that were stronger but still had some missing information, akin to the incomplete bridging described above, or more commonly, those that lacked the linguistic resources necessary to smoothly rebut the Counterargument presented.

Designating a Counterargument/Evaluation pairing as Successful meant that there was a valid and logically expressed relationship between the two stages:

Some people say that bears should not be in zoos because they need a lot of space. I disagree because if they don't have some space they can build a new cage for the bears.

(Sample 107, 1-2/AAGL)

Some people say we should not put them in zoos because bears are 7 feet tall and zoo cages are small. I disagree because they can build a huge cages. (Sample 2, 4/AAGL)

These two responses are representative of the students—nearly half the class—who were able to successfully present and evaluate a Counterargument. The Evaluation clearly refers back to the point raised in the Counterargument, making the writer's thought process easy to follow.

The students who performed here represented a range of reading and language proficiency levels. Sample 107 comes from a student who tested at ELPA Level 2 and was in the “At or Above Grade Level” reading group, while the writer of Sample 18 was in the same reading group but tested at ELPA Level 4. Neither grouping appeared to be a determining factor in performance on this part of the task. Instead, analysis showed that responses were highly sensitive to the nature of the counterargument students chose to present.

All of the students who were successful with this set of stages, regardless of their overall position, chose the same issue to discuss: that of bears' size or need for space. Students who chose to write about this in their Counterarguments were able to craft Evaluations stating their disagreement and giving a simple reason why they were not persuaded. For those writing in favor of the zoo solution, as in the examples above, this was generally some variation on the idea that the concern about bears' need for space can be addressed and overcome. Those writing against putting bears in zoos inverted this

reasoning, using their Counterarguments to raise the possibility of building bigger cages but insisting in their Evaluations that the cages would never be big enough:

Some people say that they should make a bigger cage in the zoo. I disagree because if they make bigger cages they still aren't a lot of space because they need 20 Oak Groves.

(Sample 18, 3/SR)

Sample 18 shows a particularly strong Evaluation. The information introduced in its fourth clause (“because they need 20 Oak Groves”) explains why the general objection (not enough space) is important; this level of detail is one thing that distinguished Successful from Partially Successful examples. But perhaps most important here is that the Counterargument can be directly rejected. Several options for constructing a logical, straightforward Evaluation are readily apparent.

Other Counterarguments, however, do not lend themselves to such a direct response. Here are two examples characteristic of Partially Successful responses:

Some people say we should put grizzly bears in zoos because their habitat is shrinking. I disagree because the zoos are not the same size as 20 Oak Groves. (Sample 124, 4/SR)

Some people say that we can put grizzly bears in zoos, because they will be safe. I just disagree because grizzly bears don't like to be next to humans. (Sample 112, NSS/AR)

These examples both picked up on counterarguments other than that of bears' size/need for space. In Sample 124, the writer presents the counterargument that grizzly bears' habitat is shrinking due to humans' encroachment on their territory. Sample 112 picks up on the issue of bears' safety: many grizzlies are shot in the wild, either for sport or because humans are scared of them. In both cases, it is impossible, or at least unreasonable, to totally reject the counterargument: these are, unquestionably, valid

concerns. This fact poses a challenge to writers above and beyond the one faced by students who chose to talk about bears' need for space. Here the writers need linguistic tools to concede the point to some degree but present an overriding concern. Most second graders had not yet refined such skills, and as a result Evaluations in this group read as somewhat disjointed responses to the Counterarguments. The writer of Sample 112 came the closest to expressing tacit acceptance of the Counterargument with her use of "just", but couldn't or didn't elaborate.

The extent to which the facts available would shape performance on the writing task was not something we anticipated in planning or delivering the argument unit lessons. Evaluation of the factors contributing to students' struggles with this stage pair suggests that performance could be improved by adding instruction in and practice with linguistic resources for contrast and concession, a point to be discussed in detail in Chapter 6.

A Framework for Evaluation of Argument

Evaluation of students' arguments in terms of the three performance levels discussed above was useful in identifying and characterizing what writers did well and where they tended to struggle. As I continued to work with the data, however, it became clear that there were finer distinctions to be made than those afforded by these broad categories. I struggled in some cases to classify a stage pairing as either Partially Successful or Unsuccessful. To what degree, for instance, should a student be considered to have achieved the goals of these stages, for this task, if she established a clear logical link between her Counterargument and Evaluation—but both were based on a misunderstanding or misreporting of the issues? What if a student wrote his Evidence

clearly and explained why it was important, but framed it in such a way that it supported the opposite position rather than his own claim? Such considerations led to the emergence of five dimensions along which arguments varied, both across and within the broader performance levels.

The five dimensions comprising this framework are: Logic, Factual Correctness, Consistency, Function, and Features. Logic and Factual Correctness are straightforward and not particular to an SFL-based approach. They ask, simply, whether the stages being evaluated make sense and accurately reflect source material and/or classroom discussion. Function refers to whether the stage is generally recognizable—whether what is written functions to move the piece forward in a way that aligns with expectations. Consistency and Features concern the relationships between and language features characteristic of the stage pairs. Consistency in second grade is about the internal coordination of ideas: whether Evidence and Reason actually support the Claim; whether the Counterargument actually represents an opposing position; whether the Evaluation addresses the issue raised in the Counterargument. Finally, Features is a finer-grained examination of the students’ use of the features that were in focus in classroom activities; in this argument unit, this dimension would primarily concern the use of modality to make predictions and/or form conditional sentences for the Evidence/Reason stage pairing, and language used to introduce and rebut opposing views at the Counterargument/Evaluation stage. Table 4.5 presents the second-grade framework, including questions used to guide evaluation for each dimension.

Evaluative Dimension	Criteria: Evidence and Reason	Criteria: Counterargument and Evaluation
I. Logic	Does it make sense? Is the reader able to follow the argument without needing additional background?	Does it make sense? Is the reader able to follow the argument without needing additional background?
II. Factual Correctness	Do(es) the statement(s) accurately reflect the source material and/or classroom conversation? Does the writer understand (is the writer able to explain) the concepts?	Do(es) the statement(s) accurately reflect the source material and/or classroom conversation? Does the writer understand (is the writer able to explain) the concepts?
III. Function	Is the general shape of the stages discernible? Do Evidence and Reason “do” what they’re supposed to? Does the author use language to signal movement between stages?	Is the general shape of the stages discernible? Do Counterargument and Evaluation “do” what they’re supposed to? Does the author use language to signal movement between stages?
IV. Consistency	Do Evidence and Reason, in fact, support the writer’s claim? Does the Reason address the particular piece of evidence presented?	Is the Counterargument one that would be plausible for someone with a different claim to make? Is it, in fact, counter to the argument the writer is making? Does the Evaluation effectively address the counterargument presented (do the stages “match”)?
V. Features	Does the writer use language features associated with these stages where appropriate (e.g., modality associated with making predictions/conditional sentences)?	Does the writer use language features associated with these stages where appropriate (e.g., “Some people say”, “I disagree because”, “That’s true, but”)?

Table 4.5. Evaluative framework for argument (second-grade version)

None of these dimensions on its own is enough to make or break an argument, and in fact, it is rare for any student to either meet or fail to meet all of these standards. Taken together, though, they provide a snapshot of the strength of the argument across multiple levels of meaning.

While descriptions could be adapted for evaluation of the piece as a whole, this framework is not meant to be a rubric and its application should not result in a “score”. Rather, its ideal use is as a kind of formative assessment, used to assist the teacher in providing targeted individual feedback or deciding how to modify classroom instruction in response to students’ work. (It also has the potential to be modified for use as a peer feedback or self-evaluation tool.) At the second-grade level, a simple Yes or No response

for each dimension/question gives a sufficient picture of how the stage pairing might be strengthened.

The next section presents an application of the framework using the arguments of two focal students.

Focal Students

Samir. Samir was a soft-spoken, quietly engaged child with a soft voice and a bright smile. He spoke Arabic at home with his family, and tested at ELPA Level 4 shortly before the argument unit started. He was in the At Risk reading group. In recommending him as a focal student, Ms. Beydoun commented that Samir “tries. He’s always trying. Language is low for him...[but] if I were to teach...something in phonics, he’ll be the one in the reading group to use it right away, and he’ll remember to use it too.” And true to his teacher’s observations, Samir was a highly observant participant throughout the unit. He locked his attention onto the instruction and had consistently high recall of previously taught concepts. At times it was difficult to understand his speech but his questions and contributions made it clear that he was picking up the scientific and generic concepts, including metalanguage.

In addition to his in-class performance, Samir had very strong recall of the information when interviewed following the conclusion of the unit. At the beginning of the interview, when asked what we had learned about in the unit, he gave a highly detailed response including the general topic, the issue laid out in the prompt, and a list of the stages in order, with minimal help from me regarding the names of the Counterargument and Evaluation. He then identified these stages in his own writing with ease, although Evaluation did not seem to register for him as a stage separate from

Counterargument, which, as previously discussed, was likely due to the way these stages were presented in the lessons. Samir had some difficulty in terms of identifying key language features: he incorrectly identified *are* (“If mother bear and cubs are separated...”) as a “likelihood word” in his Reason and did not show an awareness of the words typically used to begin the Counterargument and Evaluation. Overall, Samir’s responses were among the stronger in the group as regards explicit knowledge of the genre structure.

Samir’s original argument is shown in Figure 4.1 below. The transcribed version, with spelling and punctuation standardized, reads as follows:

Grizzly bears are becoming endangered. There used to be 50,000 grizzly bears but now there are only 1,000 left in the U.S.A. We should not put grizzly bears in zoos. Cubs need to stay with their moms for 2 or 3 years. If mother bears and cubs are separated the cubs won’t learn to catch fish. Some people say the bears will be safe at the zoos. I disagree because cubs won’t know how to catch fish. We should not put grizzly bears in zoos. (98)

In stage analysis, both the Evidence/Reason and Counterargument/Evaluation pairings were classified as Partially Successful, primarily due to the quality of the logical relationships between stages. Evidence and Reason were on the higher end of the Partially Successful category:

Cubs need to stay with their moms for 2 or 3 years. If mother bears and cubs are separated the cubs won’t learn to catch fish.

Writing Prompt. Should people put grizzly bears in zoos? Use specific evidence and reasons to explain or why not.

Grizzly bears are becoming
endangered. There used to be
50,000 grizzly bears but now
there are only 1,000 left in
the USA. We should not put grizzly
bears in zoos.
Cubs fight with their moms
for 2 or 3 years. If mother
bears and cubs are separated the
cubs won't learn to catch fish.
Some people say the bears
will be safe in the zoos. I
disagree because cubs don't
know how to catch fish.
We should not put grizzly bears
in zoos.

Figure 4.1. Samir's handwritten argument

These statements are both factually correct. They support Samir's Claim ("We should not put grizzly bears in zoos") and the stages take the expected forms: a fact from the text and a move to explain why it is relevant to the claim. Even the conditional sentence is properly constructed. What prevents this pair of stages from being most effective, and accordingly given the strongest possible classification, is the need to take

an additional step toward rounding out the logic behind this argument by either establishing definitively that zoos separate mother bears and their cubs or explaining why learning to catch fish would be important in an environment in which bears are supplied with all the food they need. The latter may be an example of a child assuming that preventing animals from doing or becoming what nature dictates is self-evidently negative, and that may well be an entirely acceptable assumption for a second grader. The former, for its part, is another example of how source material can constrain the strength of students' responses to the prompt: nothing in the Grizzly Bear Fact Sheet or other texts provided information about how likely it is that bear families would be separated in zoos, and this precludes the student's speaking from a place of factual authority. Had Samir chosen a different piece of evidence, he may well have composed a Successful pairing. As is, he did quite well with the informational resources he had at his disposal.

Samir's Counterargument and Evaluation were an example of the problem common to students who argued against putting bears in zoos:

Some people say the bears will be safe at the zoos. I disagree because cubs won't know
how to catch fish.

Having chosen to address the safety issue in his Counterargument, Samir was not in a position to use a "direct rejection" strategy with his Evaluation, and was unable to marshal the linguistic resources necessary to frame that stage as a partial concession to the point made in the Counterargument. For that reason, this pair of stages was marked as Partially Successful. Application of the evaluative framework highlights the same issue. The stages attempt to achieve the expected goals and their characteristic language features ("Some people say" and "I disagree because") are present and appropriately

used. They are factually consistent with the source material and classroom discussion. The dimensions along which they are weaker are Logic and Consistency, in this case two aspects of the same problem. Samir was unable to “match” Evaluation and Counterargument in a way that made their logical relationship clear. Seen—as it should be—in the context of the entire piece, the Evaluation is not entirely unrelated to the argument being established. The stage-to-stage consistency, however, remains in need of improvement.

While Samir’s performance was representative in some ways, each student brought a unique set of skills to the argument writing task. The next focal student, Hassan, had a different set of strengths and challenges.

Hassan. Hassan was a lively little boy with a twinkle in his eye and, frequently, a mischievous grin on his face. He was a leader of a small group of friends with whom he whispered and played whenever he had the chance. When asked why she recommended him as a focal student, Ms. Beydoun responded: “With Functional Grammar, it clicked [for] him pretty quickly; I was impressed. He’s always trying, always trying. His hand is always up, which I love about him.” Like Samir, Hassan tested at ELPA Level 4, although he spoke English at home with his parents, fluent English speakers. Ms. Beydoun described his speech as “very low” and “not grammatically correct”, and said that issues with language made it hard for him to write. Hassan was in the Some Risk reading group, a level higher than Samir. He was a frequent contributor to class discussions, offering ideas and at times picking up on information that his peers had not noticed or mentioned. (It was he who brought to the class’s attention the notion that grizzly bears need to stay with their mothers; before he raised this point during small

group work on Day 3 of instruction, it was an element of the Grizzly Bear Fact Sheet that had gone largely unnoticed.)

In his post-instructional interview, Hassan demonstrated a solid understanding of the genre structure, responding to requests to identify the stages in his writing without need for further explanation on my part (which a number of his peers did need). When asked why he chose the fact he reported in his Evidence, his answer demonstrated an awareness of purpose and audience: he picked that fact, he said, “because it proves it for the people who are saying we should put them in the zoos.” Hassan was less adept at answering questions about particular language features. He reported “if” as the only word in his Reason indicating how sure he was, failing to address the modal “will”, and responded that the Counterargument began with “some” rather than “Some people say”. These responses notwithstanding, Hassan’s interview supported the sense developed during instruction that he had a fairly strong understanding of the function of the stages. Figure 4.2 presents Hassan’s handwritten argument. It is immediately evident that, while his handwriting is still very much that of a second grader, it is significantly easier to read than was Samir’s. But surface appearance, in this case, belies another kind of problem. This is the transcription of Hassan’s argument:

Grizzly bears are becoming endangered. There used to be 50,000 grizzly bears but now there are only 1,000 left in the U.S.A. We should not put grizzly bears in zoos. Grizzly bears are 7 feet tall. If we put grizzly bears in zoos because they will get poisonous fish. Some people say we should put grizzly bears in zoos because their habitat is shrinking. I disagree because the zoos are not the same size as 20 Oak Groves. We should not put grizzly bears in zoos.

In the stage analysis, Hassan's Evidence and Reason were marked as Unsuccessful:

Grizzly bears are 7 feet tall. If we put grizzly bears in zoos because they will get poisonous fish.

In this pair of stages, the reader is first struck by the linguistic infelicity in the Reason. The juxtaposition of "if" and "because" results either in a conditional sentence lacking a main clause ("If we put bears in zoos, [effect] because they will get poisonous fish") or the writer has simply included an extraneous "because". Either way, it is actually a less serious problem in terms of developing the argument than are flaws related to factual correctness, logic, and consistency. The notion of "poisonous fish" is not a fabrication; the effect of contaminated water on salmon and other organisms in the Alaskan River Ecosystem was a central topic in the prior FG unit and had been briefly revisited in discussion at the beginning of the argument unit. Neither the source texts or classroom discussion, however, provided any reason to believe that zoos would provide grizzly bears with poisoned fish. Moreover, the discussion of poisonous fish in the Reason is wholly unrelated to the point raised in the Evidence regarding bears' size. In or out of context, this Evidence and Reason simply do not function together logically in support of the Claim.

Turning our attention to the Counterargument and Evaluation, we will see that despite his missteps with the previous stages, Hassan does a better job here:

Some people say we should put grizzly bears in zoos because their habitat is shrinking. I disagree because the zoos are not the same size as 20 Oak Groves.

Writing Prompt. Should people put grizzly bears in zoos? Use specific evidence and reasons to explain or why not.

Grizzly bears are becoming endangered. There used to be 50,000 grizzly bears but now there are only 1,000 left in USA. We should not put grizzly bears in zoo.

Grizzly bears are 7 feet tall. If we put grizzly bears in zoo because they will get poisons fish.

Some people say we should put grizzly bear in zoo because their habitat is shrinking.

I disagree because the zoos are not the same size as zoo Dearborn. We should not put grizzly bears in zoo.

Figure 4.2. Hassan's handwritten argument

This set of stages was marked Partially Successful. The fact that Hassan got different scores on the two stage pairs is not exceptional: only a minority of students received the same score, for better or for worse, on both pairs of stages. The majority were given two different but contiguous scores (i.e., Successful and Partially Successful; only two students received Successful marks on one and Unsuccessful on the other).

Once again, as a student arguing against putting grizzly bears in the zoos, Hassan had no choice but to select a Counterargument that was difficult to rebut. That grizzly

bears' habitat is shrinking is beyond dispute. And to his credit, Hassan attempted to construct an Evaluation that directly addressed the habitat loss issue by pointing out the limitations of the proposed replacement habitat. The Evaluation falls short only in that it fails to establish that grizzly bears *need* a space the size of 20 Oak Groves. Had Hassan written something to that effect in his Evidence this omission would be trivial, but in order for the argument to rise to the level of Successful it would have to be included somewhere. In terms of the evaluative framework, this is essentially a minor problem related to logic, one that could have been easily addressed if attention were called to it specifically. Otherwise, the Counterargument and Evaluation are factually correct; carry out their expected functions; demonstrate appropriate use of targeted language features; and come as close to being consistent or “matched” as possible given the previously discussed curricular limitations.

Traditional Primary Trait Assessment

In addition to the pre- and post-instruction tasks, independent raters used the primary trait rubric (following the same procedure outlined above) to score the arguments produced as a product of the instruction. Exact and adjacent agreement for these arguments in the initial round of scoring was high: 80% on Ideas & Content, 92% on both Organization and Voice, 100% on Word Choice, and 84% each for Sentence Fluency and Conventions (for additional detail, see Appendix N).

The scores provided by the independent raters support the overall stage analysis and suggest that the stage-based evaluation of arguments used for this study is compatible with more traditional forms of assessment. One example of this is difference in how arguments were scored based on the position the author took. As discussed above,

adopting the position that grizzly bears should not be placed in zoos meant taking on a level of challenge beyond that faced by those arguing in favor of putting bears in zoos and in excess of what most second graders were able to do. And indeed, in independent ratings student arguments in favor of putting grizzly bears in zoos were rated higher on average than arguments against putting bears in zoos on all six traits measured by the rubric (see Table 4.6).

	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
<i>Claim</i>						
For	4.73	4.25	4.37	3.82	3.72	3.57
Against	4.08	3.63	3.70	3.44	3.29	2.90
<i>Evidence/Reason</i>						
Successful	4.67	4.22	4.32	3.77	3.62	3.47
Partially Successful	4.17	3.61	3.80	3.47	3.33	2.88
Unsuccessful	4.00	3.79	3.54	3.50	3.42	3.21
<i>Counterargument/Evaluation</i>						
Successful	4.58	3.98	4.20	3.74	3.65	3.38
Partially Successful	4.25	3.98	3.93	3.63	3.47	3.17
Unsuccessful	3.92	3.33	3.42	3.09	2.92	2.58
<i>ELPA Level</i>						
1 & 2s (n=5)	4.83	4.40	4.57	3.97	4.00	4.03
3s (n=6)	4.20	3.78	4.03	3.31	3.33	2.97
4s (n=5)	3.87	3.37	3.10	3.30	3.07	2.60
5s (n=6)	4.45	3.89	4.03	3.75	3.53	3.17
<i>Reading Level</i>						
At/Above Grade Level (n=10)	4.80	4.42	4.53	3.95	4.00	3.85
Some Risk (n=7)	3.72	3.20	3.28	3.14	2.89	2.64
At Risk (n=10)	4.24	3.74	3.80	3.50	3.24	2.76

Table 4.6. Average scores given by independent raters to student subgroups

Comparing the average scores for students at different stage-based performance levels provides additional evidence that such an approach could be integrated with existing forms of assessment. Arguments with Successful Evidence/Reason pairings

scored, on average, 0.67 points higher on *Ideas and Content* and 0.78 points higher on *Voice* than the arguments with Unsuccessful pairings. A Partially Successful Evidence/Reason pairing accounted for an average decrease of 0.50 points on *Ideas and Content* and 0.52 points on *Voice* compared to arguments rated Successful on this point.

A similar pattern was observed for stage-based scoring of Counterargument/Evaluation pairings. Arguments receiving a Successful designation rated, on average, 0.66 points higher than Unsuccessful and 0.33 points higher than Partially Successful pairings in *Ideas and Content*. For *Voice*, arguments with Successful Counterargument/Evaluation stages were rated 0.78 points higher than those that were Unsuccessful, and 0.27 points higher than the essays with Partially Successful pairings.

As to whether these differences could be attributable to language proficiency or grade level, the independent ratings suggest otherwise. While those scoring at the highest language proficiency levels--Levels 1 and 2--were rated higher across the board than arguments written by students at any of the other ELPA levels, arguments written by students at the lowest proficiency level scored uniformly higher than the mid-range groups. So too with reading levels: students reading At or Above Grade Level outperformed the other two groups, but average scores for the At Risk group were higher on all dimensions than those given to the Some Risk group. These data are particularly encouraging. It is perhaps unsurprising that students at the highest reading and language proficiency levels outperformed their peers. Scores for students on the lower ends of both categories, however, show that the argument unit provided support strong enough to allow even students who normally struggle to be more successful than might be expected in writing the genre.

The scores given to focal students provide additional insight into how Oak Grove teachers approached the evaluation of second-grade writing. Independent raters gave Hassan’s argument a 4 in Ideas & Content; 3.5s in Organization, Voice, and Word Choice; and 3s in Sentence Fluency and Conventions. This means that his performance was interpreted to be fairly average amongst students with the same Claim and performance levels for Evidence/Reason and Counterargument/Evaluation, and above average for his reading and language proficiency groups. The only comment given by a rater—“Clear opinion but lack of details”—hints, as highlighted in the stage analysis, at the need to provide more factual information in support of the argument. By and large, however, these scores provide a reason for optimism, as they are arguably higher than one might expect for a student who was described by his teacher as having trouble writing due to language issues.

Samir’s argument was given a 4 in Ideas & Content; a 3.33 in Organization; 3s in Voice, Word Choice, and Sentence Fluency; and 1.33 in Conventions. These final numbers represent three sets of scores. Different raters gave this argument a 3, 4, and 5 on Ideas and Content; 2s, 3s, and 4s on Organization, Voice, Word Choice, and Sentence Fluency; and 1s and a 2 for Conventions. Their comments suggest that Samir’s struggles with conventions may have had an outsized influence on the lower scores. The comment left by the rater assigning the lowest scores read “Repeating same sentences, writing is difficult to understand.” The rater that gave the highest of the three sets of scores, on the other hand, wrote that the argument was “Hard to read, but if you can read it it does give a clear opinion and a concluding statement,” referring to Samir’s still-developing handwriting. This points to an area in which there is a potential, if not likely, mismatch

between the goals of the FG program and the values emphasized in traditional assessment of young students' writing. As discussed above, Functional Grammar instruction focuses attention on ideas and the way language is used to express and organize them; correction of surface-level features is seen as the last step in the writing process. This is counter to the kind of feedback we frequently observed teachers giving students in their classrooms, which tends to focus as much or more on issues of spelling, capitalization, and punctuation than the content of the writing. While this practice is understandable given that these teachers are held responsible for their students' mastery of these skills—and is clearly necessary if said skills have a strong effect on how essays are scored, especially on standardized tests—its ubiquity means that integrating the FG perspective with traditional forms of assessment requires conversation and awareness-raising around what each approach has to say about what is valued in writing. Ideally, Samir's writing would benefit from both approaches, attending first to fine-tuning his ideas and then focusing on making his writing visually accessible and comprehensible to readers.

The next chapter presents analysis of writing data collected at the fourth-grade level.

CHAPTER 5

Fourth-Grade Writing

This chapter provides context for the production of fourth graders' arguments and presents the results of writing analysis. The chapter begins with a description of the research context and the day-by-day unfolding of lessons in the argument unit. This description is followed by presentation of scores assigned to pre- and post-instruction writing by independent raters, indicating that most fourth graders wrote more highly-rated arguments after participating in the argument unit. Next, I discuss clause- and stage-level analysis of student arguments for the fourth grade class as a whole as well as for academic achievement and language proficiency subgroups, using examples to illustrate the range of performance observed in writing at this grade level. This characterization of student writing yields a number of pedagogical recommendations tailored to the specific aspects of argumentation that proved challenging for fourth graders in this study. Fourth grade modifications to the framework for evaluation of argument introduced in Chapter 4 are presented in the latter part of this chapter. Three focal student cases are used to show how the different analyses discussed throughout the chapter can come together to paint a detailed picture of the work of individuals as well as to highlight differences in performance by students in the same grade. Finally, I present the results of independent raters' primary trait scoring of the fourth grade writing, which support several key findings of my analysis and call attention to important differences in the values encoded in SFL-informed evaluation as compared to traditional assessments such as the primary

trait rubric used in this study. This chapter does not include a detailed comparison of second and fourth grade writing; for such a discussion, see Chapter 6.

An Overview of Lessons in the Argument Unit

The following section provides an overview of the instructional context in which the writing data were produced. The unit consisted of six lessons, which totaled approximately seven and a half hours of instructional time, plus additional writing time for some students. This section describes the step-by-step unfolding of those lessons.

Day 1 (1 hour, 5 minutes). The first day of the argument unit in fourth grade was devoted to setting up the problem around which students would build their arguments. Mr. J. spent the first ten minutes reviewing scientific vocabulary and concepts that had been established in the prior FG unit, including organisms, communities, environment, ecosystem, and human impact. He introduced the concept of intended and unintended consequences, illustrating the point by talking students through some everyday examples and having them share their own. Once this idea had been established, Mr. J. handed out the *Stopping the Cane Toad Invasion* text (see Appendix H). This text introduced the issue students would spend the next week exploring: in an ill-advised attempt to reduce the population of beetles that were destroying Australian sugar cane crops in the early 20th century, scientists there imported the South American Cane Toad (*Bufo marinus*), which not only did not eat the beetles but also turned out to be a highly destructive invasive species that reproduced prolifically and, due to its production of a potent toxin, had next to no natural predators.

The class spent approximately 30 minutes reading through the text, pausing between paragraphs to answer Mr. J.'s comprehension-checking questions and discuss

how the theme of intended versus unintended consequences had played out in the Australian situation. To solidify students' understanding of key facts, Mr. J. wrote two guiding questions on the interactive white board at the front of the room: *Why can't other animals kill the toads?* and *What will happen to the ecosystem [as a result of the toads not having predators]?* He first asked students to discuss these points with a partner, encouraging them to incorporate into their predictions the "likelihood words" they had practiced using in the previous unit (mostly modal verbs; a slightly expanded version of the list generated in the second-grade class), which were collected on a poster hanging in the back of the room. He then took notes on the white board as students shared their answers. The students correctly reported that the toads' skin was so poisonous that would-be predators died almost immediately after making contact, and demonstrated a solid understanding of the scientific content of the prior unit, predicting that the ecosystem into which the toads had been introduced would be thrown out of balance as predator populations decreased and prey populations increased.

The last part of this half hour segment was spent generating potential solutions to the cane toad problem. Students came up with their own ideas, talked with one another about them, and then reported back to the class, and Mr. J. facilitated a discussion of the likely benefits and drawbacks for each proposed solution. Students were highly motivated to share their ideas and were highly engaged in the discussion, which served as an additional opportunity to strengthen the class's shared understanding of the facts relevant to the cane toad problem.

Finally, a one-page text titled *Using Parasites to Control the Cane Toad Invasion*

(see Appendix I) was introduced. After Mr. J. made sure the class understood what parasites were, the class spent about 20 minutes reading and discussing the potential for lungworms to be used as a solution to the cane toad problem. As their name suggests, lungworm parasites make their way into their hosts' lungs, killing up to 30% of young toads and weakening and restricting the growth of older ones. The lungworms affecting cane toads do not cause as much harm to native Australian toads; DNA testing showed that the parasites were actually South American in origin, and thus did not evolve to live in Australian species. The text presents the lungworms as a promising solution that has yet to be tested outside of a laboratory setting. While reading the text Mr. J. again had students focus on modality, this time asking them to pay attention to how it was used in the text to get a sense of how sure scientists were that lungworms would be a viable solution. The text had been constructed in part to lend itself to such a discussion, and students were able to identify many places throughout the text that showed that lungworms were not being presented as a guaranteed solution.

Day 2: 1 hour, 10 minutes. The main focus on Day 2 was introducing the argument genre. The first 20 minutes or so were spent reviewing and expanding shared knowledge of the cane toad problem. After a brief discussion about what students remembered from the prior day's lesson, Mr. J. introduced a two-minute video about cane toads from Animal Planet's "Weird, True, & Freaky" series, with instructions to jot down anything presented in the video that was not in the Cane Toads text. After watching the video twice, students shared back the new information they'd noted, including that the toads can grow to be 9.5 inches long and up to 4 pounds; they lay 35,000 eggs twice a year; and they eat anything that can fit in their mouths including mice, insects, and

snakes. This new information helped reinforce the magnitude of the destruction cane toads wreak on the ecosystem; it would also serve as an additional source of information for students when they wrote their arguments.

After this discussion, Mr. J. transitioned into the deconstruction phase of the unit. He presented students with a model argument that took the position that lungworms were a bad solution to the cane toad problem (see Appendix J). Students took turns reading aloud as their classmates followed along, and Mr. J. asked clarifying questions to draw out the important points being made in the model. This took just less than ten minutes. Mr. J. then presented students with a description of the argument genre, using PowerPoint slides to present the overall purpose, the stages, and their functions. (This information was replicated on the Text Type Overview handout distributed to students; see Appendix K). In presenting the overall description of the genre (“To take a position on an issue, support it with evidence, and explain the reasons why”), he made clear the connection between the abstract description and its application in the model the class had just read (e.g., “What’s their position?” “Did they use evidence?”). When he moved on to display the list of stages, he asked students to think about whether any of them looked familiar. The intent here was to have students draw on their knowledge of a related FG genre—character analysis—that used some of the same stages to construct a literary argument (Moore and Schleppegrell, accepted). Students immediately recognized Claim (“...when they give you a question, a prompt, and you have to answer it,” one student explained) and Evidence (“something to prove your claim,” offered another). Students also made guesses about the purposes of stages they hadn’t seen before, and were collectively able to come quite close to correctly describing several of them. Mr. J. incorporated their

contributions into his presentation of the stages, refining students' phrasing as necessary to communicate the point clearly. This top-down, teacher-led part of the lesson was dynamic and lasted only about five minutes, just long enough to sufficiently familiarize students with the stages they would be exploring in the inductive activity that followed.

Once this basic understanding of the stages had been established, students practiced identifying them in their copies of the anti-lungworms model. They first worked by themselves or with a partner to identify one stage, and then there was whole-class discussion of their findings as a student volunteer came to the front to circle the sentence(s) he or she had identified on a projected copy of the model. In several cases there was some disagreement about where one stage ended and another began, and lively discussion ensued. As writing becomes more sophisticated, boundaries between stages become more fluid. Even a model written at about a fourth-grade level has more complex ideas, better-developed transitional language, and thus stages that are more difficult to parse than the very clearly demarcated stages in the second-grade models. As this complexity builds, thoughtful discussion around text deconstruction becomes ever more valuable.

The discussion also provided an opportunity to highlight the different kinds of meaning used to build the genre's stages. Take, for example, the paragraph containing the *Description of Problem* and *Claim*:

The cane toad invasion has been a disaster for the ecosystems in Australia. Cane toads were brought to Australia to solve a problem, but they have become an even bigger problem themselves. The huge toads eat anything that will fit in their mouths, and nothing can eat them because they are highly poisonous. *Scientists*

think that spreading the cane toad lungworm around could help solve the problem, but it is too dangerous. They should not use it.

The first student to come to the front of the room, Yasmin, identified the first three sentences as the Description of Problem. Several students immediately disagreed, and another student, Amir, stood up to say that he had also included the fourth sentence, italicized above, marking only the last sentence as the Claim. Yasmin's response reflected the research team's thinking in designing the materials, but Mr. J. took Amir's point and turned it over to the class for consideration. The Description of Problem stage had been characterized as including the factual information necessary to contextualize the argument for the reader; Claim, in contrast, was about the writer's opinion. Mr. J. asked students which kind of meaning the sentence included. By momentarily narrowing his focus in this way, he was pushing students' thinking regarding the boundaries between stages. The sentence in question contains both factual information ("Scientists think that spreading the cane toad lungworm around could help solve the problem") and the writer's opinion ("but it is too dangerous"): it is doing the work characteristic of both stages. On the projected model, Mr. J. ultimately marked this sentence as part of the Claim, but took an additional step toward acknowledging the ambiguity by underlining the last clause in the fourth sentence along with the fifth sentence, summarizing his and the students' reasoning as he did so:

"I'm going to circle all of this as part of the Claim, but I'm going to underline in dark—which parts are really the most opinion? This is the heart of the Claim, right? 'It is too dangerous, they should not use it.' That is really what the Claim is. So I can understand why you thought that that first part could be part of the

Problem and describing the solution, because that first part of the sentence is describing, right? It's describing what scientists are considering the solution... So you could [take either Yasmin's or Amir's approach.] Sometimes it's not completely clear-cut. But in those sentences we really get the opinion.”

This kind of discussion is a product of a curriculum designed to weave together the explicit guidance offered by the teacher with opportunities for inductive learning. The presentation of the genre could easily have taken the form of Mr. J. simply reading the stage descriptions and pointing out where they occurred in the model. Instead, students were asked to interact with the text in a much deeper way, reasoning through why a given sentence should be labeled as one stage or another.

Another productive student-initiated discussion came up almost immediately after the previous example, as the class moved on to identify Evidence and Reason in the model. This model included Evidence and Reason pairs. The students quickly found and labeled the first pair, which featured fairly straightforward examples of the stages. The second pair, however, featured a common linguistic red herring—introducing Evidence by calling it a reason—that a student named Mohammad was quick to point out:

Mohammad: Mr. J., it says “another reason this is a bad solution...”¹⁸

Mr. J.: That's tricky. You have to think about what it's doing. That word “reason” can be very tricky. Now, the Reason tells us why and how the Evidence supports [the Claim]. Does this tell us exactly why and how? ‘Another reason this is a bad idea’—does it say why it's a bad idea?

Mohammad: No.

¹⁸ Text: Another reason this is a bad solution is that the worms only kill about 30% of baby toads. That means 70% of toads would still survive. The impact is too small to place the ecosystem in such risk.

Mr. J.: No. But I like that you found that, Isa. Because I knew that it could be tricky. But what is it doing? It's introducing the evidence. The evidence is a fact that supports it, that only 30% are killed. Now, 70% would still survive—is that a fact?

Several students: Yes.

Mr. J.: It is! So should we consider that part of Evidence, or Reason?

Several students: Evidence.

Mohammad's observation provided Mr. J. with another opportunity to help students think about the type of meaning being presented in a particular piece of text, which he describes here as “think[ing] about what it's doing.” Consistently emphasizing meaning and function was a central goal for the FG team in planning and implementing instruction. The Description of Problem and Claim example above showed students that the stages are not always as distinct as the genre description may imply. The example raised by Mohammad reinforced the importance of going beyond the meaning of words themselves, because words can sometimes “mean” one thing but “do” another.

Other important discussions were not planned, but came about as students began to get a sense of how stages related to one another. For example, after concluding the discussion of Evidence and Reason, the class moved on to Counterargument and Evaluation. As Mr. J. began to ask students about how they had labeled the relevant paragraph, a particularly energetic student named Isa excitedly interrupted him with an astute observation:

Isa: Isn't the last Evidence and Reason the Counterargument?

Mr. J.: Exactly. This is the Counterargument, ‘One reason to use lungworms is that they do make older toads smaller and slower.’ This is evidence—it’s a fact—but it’s in support of the other claim, it’s not in support of their claim.

Having confirmed Isa’s observation, Mr. J. continued. He started to ask about the use of modality as a means of identifying the Evaluation, but instead Jamal broke in, picking up on Isa’s point:

Mr. J.: [reading from text] ‘But smaller toads still eat a lot of food, and slower toads are still poisonous. In the end, it is highly unlikely that lungworms would make a real difference.’ Is the author really sure?

Jamal: It’s a Reason.

Mr. J.: Exactly! But it’s a specific type of Reason. You’re absolutely right. It’s explaining why that argument isn’t right. It’s an evaluation of that Counterargument. But it is a reason...it’s your explanation about why you disagree with that person.”

Drawing this parallel between Evidence/Reason and Counterargument/Evaluation was not part of the argument curriculum. Isa, Jamal, and others noticed the pattern on their own. The fact that they did points strongly to their having adopted a meaning-based approach to analyzing the text: they were able to recognize stages that, while presented as totally distinct from one another, were fundamentally “doing” the same thing. Mr. J., for his part, recognized this as an important contribution and brought it into the discussion, reiterating the ways in which the stages were similar while also highlighting their key differences.

Finally, discussion returned to the question of modality—not at Mr. J.’s direction, but because a student raised her hand to comment on the usage of “highly unlikely” in the Evaluation. The class talked briefly about how sure the author was; whether credible writers would sound more sure about Evidence or Reason; and whether it was possible to be entirely sure about what one wrote in the latter stage. This discussion brought the deconstruction of the model to about 35 minutes, at which point the lesson had to conclude for the day.

Day 3 (1 hour). Mr. J. started the third lesson of the argument unit with a five-minute review of the model argument read the day before, asking students, stage by stage, about why the author had been against using lungworms to address the cane toad problem. He then introduced a second model argument, this time one that argued in favor of using lungworms (see Appendix J), and had students work in small groups to identify the stages (i.e., one group looked for the Description of Problem and Claim, another looked for Evidence and Reason, and so on). While they worked, Mr. J. and Ms. K. visited different groups, listening and offering guidance where needed. About five minutes later, once they had completed this task, students reported their answers to the whole group.

As the first group shared what they had labeled Description of Problem and Claim, Mr. J. introduced a refined notion of the Description of Problem stage. He told students that this stage might be better thought of as describing an “issue” rather than a problem, because in order to give readers the information they needed, writers would have to not just present the problem but also introduce the solution they would be evaluating. The rest of this segment of the lesson, which lasted about 20 minutes, was

similar to the deconstruction segment of Day 2. Mr. J. and the students talked through stage boundaries, what function different sentences were serving, and how the author used “likelihood words” in the Reason and Evaluation stages to communicate how sure they were or how likely something was to happen. The purpose in drawing students’ attention to this language feature was that in most cases, using modality to hedge their statements would be a way of increasing the credibility of their arguments, as it would demonstrate an understanding of the history of the cane toad problem (which was itself an unintended consequence of the attempt to eradicate the cane beetles) and the current state of scientific understanding (even scientists that study ways to eliminate cane toads are not entirely sure about the effects that the proposed solutions might have on the ecosystem).

At this point a new text was introduced. The next twenty minutes of class time were devoted to reading and discussing the *Meat Ants* text (see Appendix L), a one-page description of another potential solution to the cane toad problem. The meat ant is an aggressive ant species that sometimes eats animal flesh and is not affected by cane toads’ poison. Despite its name, the meat ant has a diet that normally does not consist primarily of meat, but of honeydew from caterpillars. In their natural habitat, this is a mutually beneficial relationship, as the ants provide protection for the caterpillars. Native toad species that have evolved in the same environment know to flee from meat ants, but cane toads do not recognize them as a predator and sit still while attacked. Preliminary studies suggest that meat ants would be a safe solution, but this theory has not been tested on a larger scale in the wild.

Mr. J. introduced the text by telling the students that they would be writing their own arguments evaluating meat ants as a solution, and that they would be able to choose from three positions: that meat ants were a good option and should be used immediately; that they were a good option but needed to be studied further before being used as a solution; or that they were a bad solution that should not be used. Students were asked to think about what could be used as evidence for each of these positions as they read. Volunteers read paragraphs aloud, and Mr. J. asked clarifying and comprehension-checking questions at key points and took some initial notes on facts about the meat ants as they came up (at this point without noting which position they might support). He also asked students to identify instances of modality related to likelihood and judge the degree of likelihood the author was communicating. For example, a student identified the phrase “will likely” in the sentence “Getting the ants to move to cane toads' ponds will likely lower the number of ants in other places,” and decided that this meant the author thought the result was not just possible, but probable.

When the discussion of the Meat Ants text concluded, each student received a blank copy of an evidence chart handout (see Appendix D). At the top of the evidence chart was the prompt to which students would write:

Scientists in Australia think using meat ants might be one effective way to stop cane toads from spreading and hurting more organisms. What do you think of this option? Is it a good option Australia should act on now, a good option they should study more, or not a good option? Why?

Below, there was a wide column for students to take notes on evidence from the text, to the right of which were three smaller columns representing the three possible

claims; as they filled in the first column, students were to evaluate which claim(s) a given piece of evidence could support and mark the columns accordingly. This chart helped concretize a broad understanding of the potential benefits and drawbacks of using meat ants to control the cane toad population, and was intended to serve as a support for students to draw on when they wrote their Evidence.

In the just under ten minutes remaining in the class period, Mr. J. and the students worked together on a couple of initial examples. Students thought about and discussed as a class the evidence they thought was most important: that toxins from the toad are not likely to affect the ants, and meat ants can kill toads. Mr. J. filled in a projected copy of the chart with paraphrased evidence, but emphasis was placed on the importance of retaining the degree of likelihood originally conveyed by the author, as changing that meaning would alter the accuracy of the evidence. Students decided in class discussion that these facts would support using the ants, either immediately or after further study. Having completed the modeling of the evidence chart, Mr. J. concluded the lesson.

Day 4 (1 hour, 20 minutes). Day 4 began with a review of what students knew about cane toads, lungworms, and meat ants. Mr. J. quickly moved into a continuation of the collaborative work on the evidence chart, which was projected onto the interactive white board at the front of the room. To strengthen the scaffolding of evidence, he walked students through adding an explanation to the evidence put on the chart the day before, eliciting from them the “why or how” of the evidence—the mechanism by which meat ants are able to kill cane toads (the toads, used to letting their toxins do the work of fighting off predators, just sit there as they are attacked).

After expanding the first piece of evidence, the class moved on to look for additional examples. In talking through the new additions to the chart, Mr. J. began to have students practice formulating Reasons:

Isa: [reading from text] ‘Getting the ants to move to cane toads’ ponds would likely lower the number of ants in other places. This could change the behavior of other organisms.’

Mr. J.: Good. And that connects with an idea earlier in the text—

Isa: —it’s changing the behavior of the other organisms, that means it’s changing the behavior of the whole ecosystem.

Mr. J.: Okay, so...[writing on overhead] ‘There will likely be fewer ants where they normally live...This could change the behavior of other organisms.’ Now, talk through your reasoning. Why might this support not using it or studying it more?

Isa: I think it supports not using it...because if it changes the behavior of the organisms that means it’s changing the behavior of the whole ecosystem, like let’s say—

Mr. J.: —Good, so give us an example of how changing the whole ecosystem could be bad.

Isa: Like let’s say there’s a farmer and they’re growing crops and there’s meat ants, there’s like small little animals the meat ants could eat, like they usually eat them to protect the crops.

Mr. J.: And if you take them away?

Isa: There will be less meat ants, and the other animals, some of them will be eaten but mostly a *lot* of the crops will be eaten.

Mr. J., to class: So he's giving us an example, we don't know for sure, but it's possible—he's giving us our Reason for why changing an ecosystem could be bad...that's a very good example for a good Reason for that.

Isa's early comment—that changing the behavior of one organism would ultimately affect the entire ecosystem—reflected a good understanding of the science content in the argument lessons and the prior FG unit. Mr. J.'s questions scaffolded the translation of this knowledge into an explanation that was explicit enough to make Isa's thinking accessible to others. Isa's reasoning turns out to be mostly conjecture, but the exercise is nonetheless likely to have helped Isa and his classmates become more attuned to the process of building a Reason, a process further reinforced by Mr. J. guiding students offering subsequent examples through the same kind of oral rehearsal of their Reasons.

Approximately half an hour into the lesson, Mr. J. transitioned the class into individual work, asking students to think about the position they wanted to take and look for one or two pieces of evidence that supported it. Students worked on this for about 10 minutes, many discussing their findings with partners. Then, they returned to whole-group discussion, taking 20 minutes to discuss the new pieces of evidence and why they supported the different claims.

Finally, Mr. J. handed out the Evidence/Reason scaffolds (see Appendix D). As discussed in Chapter 3, one of the guiding principles in designing the argument unit was to keep the content cognitively manageable. With that principle in mind, separate

scaffolds were created for each stage or pair of stages, allowing students to write drafts of the stages one or two at a time and then transfer or modify these smaller pieces as they wrote their final argument. In handing out the Evidence/Reason scaffolds, Mr. J. asked students to think about which piece of evidence they felt offered strongest support for their position and cautioned them to pay attention to how much additional explanation might be required:

The next step, I want you to take the piece of evidence that you think is the strongest, and in a complete sentence, I want you to write down the Evidence that supports that claim. Remember, you want to include enough information so that it makes sense. Some of these you really have to explain more than others. Which ones do you have to explain more than others? Which ones really require some talking through what actually happens?

At this point students raised their hands to offer thoughts on which of the pieces of evidence displayed on the chart would require more explanation and which were more self-evident. This aspect of writing—anticipating just how much information the reader needs—is difficult to address within the confines of a single, rather short unit. By returning several times to the issue of how much detail would be appropriate, Mr. J. was doing what he could to support students in making this kind of judgment. This is ultimately a skill that continues to develop throughout the school years, and the fourth graders in this class varied widely in terms of the sophistication of their audience awareness.

Students worked on their Evidence/Reason scaffolds for 15 minutes, during which Mr. J. and Ms. K. spoke with students individually on an as-needed basis. The three

students who asked Ms. K. for help had started to fill out their scaffolds in a way that suggested that they did not yet have a clear understanding of stage boundaries (for example, inserting their Claim into the Evidence or presenting new evidence in their Reason). Others, however, seemed to understand quite well, including the student whose clearly-written Evidence and Reason draft was the focus of discussion during the last five minutes or so of the lesson. A number of other students also wanted to share their work, and were told that they would be able to the next day.

Day 5 (1 hour, 20 minutes). Day 5, like prior lessons in the unit, began with review: first of cane toad facts, and then of the structure of the genre. Mr. J. distributed a handout called *Organizing Your Argument*, a simplified version of the Text Type Overview handout from Lesson 2, and read through it, telling students he realized they had spent time thinking about the stages of the model arguments but that he wanted them to see again how it all fit together.

The class then returned to the point at which they had stopped the day before. Student volunteers came to the front, put their Evidence/Reason scaffold on the projector, and read what they'd written. Very quickly, a point that Mr. J. had tried to make in the last few minutes of Lesson 4 resurfaced. The student who had presented her example the day before was making a claim of “good, study more” (this phrase had by now become established shorthand in classroom discussion of Claims, along with “good, act now” and “bad, don't use”). In discussing her Evidence and Reason, Mr. J. started to make the point that it was possible to have two sets of Evidence and Reason, one pair showing why meat ants were a good solution, and another showing why they nonetheless needed to be studied more. Now, in Lesson 5, he returned to this point. A student volunteer presented

Evidence that supported the “good option” part of the claim, but her Reason pointed to “study more”. Mr. J. asked students to think about a piece of evidence that might be a better match for the Reason. Students had a hard time with this; when Mr. J. asked them if they understood what he was saying, they said they didn’t. He tried again, this time saying that if they took the “good, study more” claim they’d first need to prove that it was a good idea, and then that scientists should study more. He attempted to clarify the point by asking for an example that had taken the position that meat ants were a good solution—a claim which would only require the writer to write one Evidence and Reason pair—but nobody identified themselves as having taken that claim.

This somewhat dissatisfying exchange was illustrative of the experience of piloting an instructional unit. The lessons had not been taught before, and as Mr. J. delivered them he had new insights into how the writing might be best structured. It is not clear in this instance whether the students’ confusion was a result of an explanation that could have been improved with additional time to prepare, or if it is an indication that something about what was being described was too challenging to incorporate into the initial presentation of the genre at this grade level. Certainly a more condensed approach is possible, in which support for both the “good” and “study more” aspects could be established within one Evidence and Reason pair (for example, by constructing the Evidence to present a fact suggesting that meat ants would be a good solution and then noting in the Reason that scientists are not sure that certain negative outcomes could be avoided and thus the ants should be studied further before they are used as a large-scale solution). This, in fact, was the approach taken by many students in writing their first drafts of Evidence and Reason. If the difficulty Mr. J. had in communicating his point

was indeed indicative of our having departed from the principle of cognitive manageability, perhaps it would be advisable to avoid introducing this more comprehensive type of response to the prompt until after students had been familiarized with the genre. It is also possible that, had the unit unfolded over a longer period of time, any challenge posed by structuring stages in this way would have been surmountable. At any rate, as will be discussed below, few students opted to include more than one Evidence and Reason pair in their final arguments.

Students were given about 10 minutes to finish writing up their first piece of Evidence or, if they had done so the day before, to find another one. Mr. J. and Ms. K. again talked to individual students as they worked. Mr. J. gave feedback focused on choosing the best piece of evidence to support a particular claim. Ms. K.'s conversations, mostly about helping with Reasons, were similar: they focused on students convincing her or convincing their reader that the evidence they chose really showed why meat ants are a good solution. In some cases, when students were unable to explain the connection, this line of questioning led to the conclusion that evidence didn't really work (for example, the fact that meat ants are a dominant species did come directly from one of the texts, but it was not useful in supporting any of the possible claims). After this individual writing time, the class spent 15 minutes discussing student examples. In total, five children read what they had written. In this segment Mr. J. focused on helping students talk through and clarify their thinking: by adding detail; by using modality to "tone down" predictions or be more factually precise; or simply by orally rehearsing their reasoning.

The class then transitioned into a brief review of Counterargument. Mr. J. asked a student to explain the role the stage plays in the overall argument. He then displayed the Counterargument/Evaluation scaffold (see Appendix D) on the overhead projector. This scaffold was designed to help students identify a counterargument and then reflect on which part(s) of that counterargument they disagreed with, which part(s) they agreed with, and how strongly they felt about each of these. To model its use, Mr. J. asked students to come up with Evidence and Reason in support of the claim that meat ants were a bad solution (in other words, to pretend that this was the position they were taking) and had them practice identifying a viable counterargument. He used this information to talk through an example with a student, recording the student's responses on the projected scaffold. As he mediated the specific example chosen by the student, Mr. J. made several more general points that would help students think about these stages in their own writing. Because Evidence is factual, he said, any disagreement they had would likely be with the way the evidence was framed in the Reason rather than with the Evidence itself. He also modeled acknowledgment of the parts of the Counterargument that were reasonable, encouraging students to temper their language in their Evaluations to reflect that there was some validity in the opposing position.

This latter point—that a measured consideration of the facts is desirable rather than something to be avoided—is a mark of more sophisticated argumentation. It is a value that is explicitly encoded in the argument curriculum. The Counterargument/Evaluation scaffold intentionally devotes equal space to considering what one agrees with in addition to what one disagrees with. The “good, study more” option itself was included in the prompt because an intermediate approach was seen as a

very likely conclusion to reach given the facts in the cane toad situation. This is, of course, an atypical approach in elementary schools, which tend to teach persuasive writing that asks students to take a strong stance in supporting a claim. When Mr. J. asked students why they thought it would be important to include the Counterargument as part of their overall argument, their responses reflected the goals of persuasion: they thought that including a Counterargument would help persuade the reader or make someone who took an opposing viewpoint change their mind. Mr. J. offered a different idea:

Maybe there are points that you can agree with. And maybe there are points that they can agree on with you. It isn't always that we are completely against each other. They may agree on some things and you want to find a solution that works. In other words, Mr. J. was presenting a vision of argumentation that wasn't just about winning, and beginning to frame the goals for writing as something larger than the classroom assignment.

Once modeling of the Counterargument/Evaluation scaffold was complete, Mr. J. gave students about 15 minutes to work on their own. He and Ms. K. talked to students individually to help them pinpoint counterarguments based on their claims. Ms. K.'s approach in helping students understand this was to personify the opposing position. She asked for students' claims and then asked them to think of her as someone who took a different claim: what would she say to support her position? How would they respond? After this, Mr. J. transitioned into whole-class discussion of four student examples. As the students came to the front to project their scaffolds, he asked a set of guiding questions: *What is your Claim? What Claim would the person presenting the*

Counterargument make? What is the Counterargument (Evidence and Reason)? Why does this support [the other Claim]? What did you agree with? What did you disagree with? These questions varied a bit by student, and some needed more help from classmates than others, but in general the questions supported the pedagogical practice established throughout the unit of providing a process to help students make their thinking explicit. This discussion lasted about 10 minutes, through the end of class.

Day 6 (1 hour, 35 minutes; additional writing time for some students). Day 6 was the final lesson in the argument unit. This lesson was taught by Ms. K., who began with a 15-minute review of what the class knew about cane toads; why meat ants might be good or bad; what the prompt and options were; and what Counterargument and Evaluation consisted of. Students were then given 20 minutes to finish writing filling out their Counterargument/Evaluation scaffolds, as many had not finished on Day 5.

After this writing time, the class transitioned into a discussion of the Description of Problem and Claim stages. Ms. K. projected the relevant portions of both of the lungworms argument models in order to discuss “what we notice about what’s going on in the paragraphs.” Students read the paragraphs out loud, identified the two stages in both of the models, and responded to Ms. K.’s questions about the stages. As Mr. J. had done with Evidence, Ms. K. drew students’ attention to the “grain size” issue inherent to writing the Description of Problem:

Ms. K.: [reads text from model aloud] Are those the big points, are those the main ideas, or are those tiny details? ... Main idea, I hear lots of voices saying main idea. This is the ‘long story short’. This is—if you give this to...somebody in the class across the hall, they don’t know what cane toads are. ... Somebody who

doesn't know what cane toads are, somebody who doesn't know what meat ants are—they need a little bit of information to know what they're going to be reading about.

This point could have been strengthened by reviewing the source texts for specific examples of what would be considered a “main idea” versus a “tiny detail”. At this point, however, the unit had been extended by one day beyond what the classroom teacher had originally agreed to, and instructional time was at a premium, with just ten minutes allotted for the discussion of these two stages.

Ms. K. moved on to discuss the Claim stage, focusing on the strength of the Claims in the models. The class read the examples and discussed how strong the authors felt about the positions they were taking, deciding after examining the authors' language choices that one was more absolute (“They should not use it”) and the other left room for discussion (“It is worth trying”). Ms. K. asked students to be conscious of this in their own writing:

As you're writing your Evidence and you're doing your Counterargument, you probably have an idea in your head about how strong you want to make your claim, right? So when you write your Claim, I want you to think about this too. Think about, are you going to make a really strong claim? Or are you going to make a claim that has room for maybes or mights or probablys? Think about, when somebody reads your Claim, how strong they're going to think that you feel.

The overall point here was that Claim, like Reason and Evaluation, was a stage in which language choices are especially important. This group of stages needs to be

thought of a bit differently than Description of Problem, Evidence, and Counterargument, the construction of which is largely a matter of selecting the facts that best support one's claim and deciding how much the reader needs to know. It is certainly common for sophisticated writers to shape readers' perception of facts by choosing language that casts them in a particular light (for an extensive discussion of linguistic resources used to do this, see Martin & White, 2005). In this fourth-grade curriculum, however, students were asked to maintain the meanings from the original text in their presentation of facts, reserving their most careful thought about subtle differences in word choice for the stages that reflected their own evaluation of those facts. While the shorthand versions of the potential positions had been used conversationally throughout the unit, students were encouraged at this point to modify that wording so that their written Claims would accurately reflect their opinions.

After discussing the Description of Problem and Claim, the class moved on to a five-minute review of the overall purpose and structure of the genre. When asked about the purpose of writing the argument, some students volunteered more abstract reasons for writing ("To show your reader what you know about meat ants and if it's a good solution"), while others thought of their goals in more concrete terms (helping Australian scientists and/or government officials make a decision). Ms. K. picked up on this latter idea as an opportunity to increase relevance and authenticity by injecting into the unit some sense of a larger purpose for writing, and incorporated it into the subsequent review of the argument stages and how they fit together.

Finally, students were given time to write their Description of Problem and Claim, and bring all of the stages together to build their final arguments. This block of writing

time lasted for about 45 minutes, during which Ms. K. consulted with students who needed help. Some students finished writing during this block, while others needed as much as a half hour of additional time, which their teacher granted them later that day. Students who needed additional time worked independently while Ms. K. conducted interviews with focal students elsewhere in the building.

Independent Scoring of Pre- and Post-Instruction Writing

As described in Chapter 4, this study's design included pre- and post-instruction argumentative writing tasks intended to gauge students' unscaffolded performance on a prompt similar to the one used in the argument unit before and after participating in the unit's lessons. The prompt was the same for the pre- and post-instruction tasks, and the same prompt was used in second and fourth grades. As described in Chapter 4, the subject matter was drawn from the FG unit that students at both grade levels had just completed. The prompt students were asked to respond to in the pre- and post-instruction tasks read:

In Florida, there are many hotels along the sandy beach. Every morning, the people who work at these hotels go out and remove the beach wrack so the hotel guests can swim on a clean beach. Should the hotel workers take away the beach wrack? Use specific evidence and reasons to explain why or why not.

The pre- and post-instruction prompts were administered by classroom teachers, who followed instructions provided by the research team. Arguments produced in response were evaluated by independent raters who applied a primary trait rubric following the procedure described in detail in Chapter 3.

Inter-rater agreement. To determine reliability for independent scoring,

percentages of exact and adjacent agreement (defined as scores within one point of one another on the rubric’s six point scale) were calculated. Combined agreement ranged from 71% to 99% for the pre-instruction writing and from 78% to 89% for the post-instruction writing. Table 5.1 displays the combined percentages for exact and adjacent agreement; for a breakdown of exact versus adjacent agreement, see Appendix N.

	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
Pre	96%	99%	71%	78%	89%	71%
Post	85%	81%	81%	89%	78%	82%

Table 5.1. Inter-rater agreement (exact and adjacent) for fourth-grade pre- and post-tests

Surprisingly, agreement on Ideas & Content and Organization was lower for writing produced after the argument unit, though overall agreement for these traits remained high. As described in Chapter 3, arguments for which scores on any trait differed by more than one point were referred to a third rater. A review of the third set of scores for cases where exact or adjacent agreement fell below 30% for either pre- or post-instruction writing—in this case, Voice and Word Choice—found that in the overwhelming majority of cases, the third rater’s scores were within one point of those of one of the two original raters. Of the 12 pre-tests and 13 post-tests referred to a third rater, there were no cases in which there was a discrepancy of more than one point between the third score and the initial two. The “tie-breaker” round of scoring, then, tended to reinforce the scores produced in the initial round of scoring and functioned to provide additional data to round out the final scores used in analysis, which were an average for each trait of all the ratings assigned to a particular piece of writing. The next section presents a discussion of trends in these final scores.

Results of scoring. Of the fourth graders in the study, 21 students, or 80.77% of the class, improved on some or all of traits measured by the independent raters’ rubric. The

five students whose performance decreased represented a range of language proficiency and academic performance subgroups. There were two students in this group who read below grade level, one who was “approaching” grade level, and even two who read above level; likewise, three students tested at ELPA Level 3 while the other two were on monitored status. It is possible that these students were distracted during the post-instruction task, or that they took it less seriously, having completed an identical task just a week or so earlier.

Of the 21 students whose performance improved, on the other hand, 13 were given higher scores on all traits. Eight improved on some traits but their scores decreased for one or more others. Seven of these students saw a decrease in their Conventions score; for two of them, this was the only trait for which they received lower scores. Four students were rated lower in Sentence Fluency. There were only one or two instances of lower scores for each of the remaining four traits. That decreases were concentrated in the Conventions category is consistent with Functional Grammar’s emphasis on ideas and how they relate to one another over surface-level features, and is likely reflective of students’ focus being allotted accordingly; see Chapter 6 for additional discussion of this point. The appearance of Sentence Fluency as an affected category could be similarly reflective of a focus on including the elements of the genre that had been introduced in the argument unit, to the detriment of the kind of fluidity measured by this trait—albeit for just four students.

There did not appear to be either “ceiling” or “floor” effects in the fourth grade data. The students who performed best on the pre-instruction task—those with who generally had scores of 4-4.5 across traits—were essentially equally likely to receive

higher or lower scores; while some of these initially high performers did see a decline in performance on the post-instruction task, others with similar initial scores were rated even higher on the second task. The lowest-performing fourth graders had scores in the 1-2 range across traits, meaning that in a very few cases it would not have been possible for scores to decline. This was not an issue, however: children who initially scored among the lowest in the class uniformly showed improvement.

In addition, as with the second-grade class, in several cases it was the students in the lowest language and reading groups whose writing improved most in terms of the traits that aligned most strongly with the goals of the instruction. In the fourth grade, students at the Approaching and Below Grade Level reading groups showed the strongest gains for Ideas & Content and Organization. Table 5.2 shows the average change in points for fourth-grade subgroups. The Below Grade Level group in particular showed gains that nearly doubled the class average, while the Above Grade Level group showed the least improvement. At this grade level, however, there was no corresponding increase in scores for the lowest language group. Likewise, while the lowest language group's Convention scores fared worst, no such pattern was apparent for the reading groups. (For additional detailed discussion of these results, see Chapter 6; for individual students' scores, see Appendix O.)

While this study's design limits the claims that can be made regarding the effectiveness of the instruction itself, the primary trait scores provided by independent raters demonstrate that using a form of assessment that is typical in their context, Oak Grove teachers rated most fourth grade writing produced after participation in the argument unit higher on most traits than the writing produced before the instruction, with

the strongest gains typically observed for the dimensions emphasized by the FG approach. The following section reports on my own evaluation of student writing using an SFL approach to analysis.

	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
<i>Overall</i>	0.28	0.28	0.27	0.18	0.20	-0.06
<i>ELPA Level</i>						
Monitored (n=12)	0.25	0.26	0.22	0.07	0.43	0.14
1s & 2s (n=5)	0.57	0.67	0.83	0.53	0.30	0.20
3s (n=9)	0.17	0.10	0.03	0.13	-0.12	-0.42
<i>Reading Level</i>	0.00	-0.07	-0.17	-0.07	0.23	-0.07
Above Grade Level (n=5)	0.24	0.28	0.33	0.20	0.31	0.24
On Level (n=9)	0.46	0.34	0.46	0.00	-0.04	-0.63
Approaching Level (n=4)	0.50	0.56	0.46	0.48	0.33	0.02

Table 5.2. Average pre- to post-test changes in scores by subgroup, 4th grade

Written Arguments

After the unit had concluded, students' arguments (n=27)¹⁹ were transcribed and standardized to prepare them for use in linguistic analysis. Several analyses were performed in order to characterize the clause- and stage-level features of students' argumentative writing. The results were further analyzed to produce descriptive statistics for the performance of the class as a whole as well as for language proficiency and academic performance subgroups.

In Ms. Haddad's fourth-grade class, 12 students had exited ELL program services ("monitored status"),²⁰ and 10 students tested at ELPA Level 3 (High Intermediate).

Three students were at ELPA Level 2 (Proficient) and two tested at Level 1 (Advanced Proficient). To facilitate meaningful comparison, the five students testing at Levels 1 and

¹⁹ For statistical analysis one writing sample was excluded as an outlier. This sample, written by a student who scored at the lowest ELPA and reading levels, was significantly longer and less coherent than other writing collected for the study.

²⁰ These students had all tested at ELPA Level 1 as of the Spring 2011 administration of that test.

2 were grouped together for statistical calculations. Ms. Haddad also grouped students by reading level. These groups were labeled Above Grade Level (five students), On Level (nine students), Approaching Grade Level (four students), and Below Level (nine students). These groups, it is worth mentioning, overlapped to a great extent: all monitored students were reading either Above or On Level, while all 3s were in the Below or Approaching Grade Level groups (1s and 2s were split between Below and On).

Trends: Clause-level analysis. Arguments were first parsed into component clauses and T-units, and any embedded clauses were identified. A T-unit (or minimal terminable unit) is a main clause and any attached subordinate clauses (Hunt, 1965). The T-unit, which provides an approximation of syntactic complexity, is a widely-used measure in linguists' studies of language and writing development; including counts for T-units and individual clauses allows for this work to be in dialogue with SFL-oriented research as well as work grounded in other perspectives. The relationship between clauses and T-units can take a number of forms. For example, the T-unit was conceived of as the minimal unit that could be considered a complete sentence. Many sentences can be analyzed as a single T-unit and a single independent clause, with no embeddings:

Scientists have found a solution.

This, of course, is not always the case. Consider the following sentence, which consists of three separate clauses (boundaries marked with slashes), which together form a single T-unit:

Cane toads breed a lot / eat a lot / and kill a lot of predators with their poison.

The next sentence also represents one T-unit, but includes multiple embedded clauses (marked with brackets):

One reason [why meat ants are most likely a horrible solution] is [because meat ants mostly eat honeydew from caterpillars and butterflies].^{21 22}

The mean number of total clauses for the fourth-grade arguments was 28.35 (ranging from 14 to 45, with a standard deviation (SD) of 9.56). The mean number of T-units was 14.35 (ranging from 6 to 21, SD 4.85), with a mean T-unit length of 2.00 clauses (from 1.53 to 2.80, SD 0.20). All students used at least two embedded clauses, and some used as many as 13; the mean number of total embeddings per argument was 6.50 (ranging from 2 to 13, SD 3.17) with an average of 0.47 embedded clauses per T-unit (ranging from 0.15 to 1.17, SD 0.23).

Only minor differences were observed between students at different language proficiency or academic achievement levels. In terms of the latter, the only observed trend was that students in the Above Grade Level (AGL) group outperformed all other groups on all clause analysis measures except for embedded clauses per T-unit. The total number of clauses for the AGL group averaged 37.00 (range: 24 to 45), while average number of clauses for the On Level (OL) group was 25.00 (range: 19 to 33). The average total number of T-units for the AGL group was 17.40 (range: 13 to 21), compared to the next highest-scoring group, the Approaching Grade Level (APP) students, who averaged 14.25 T-units (range: 6 to 20). The average total numbers of embedded clauses for these groups were, respectively, 8.20 (range: 2 to 13) and 6.50 (range: 4 to 8). A similar, though less marked, trend was observed in terms of language proficiency groups for the

²¹ Taken from Sample 116.

²² Notation for this analysis uses brackets to mark embedded clauses and slashes to mark all other clause boundaries.

total number of clauses measure: the monitored group used an average of 30.25 total clauses, while ELLs at Levels 1 and 2 averaged 29.00 clauses and those at Level 3 used an average of 25.44.

While one might also expect to observe increased complexity in the form of increased embeddings and clauses per T-unit (Schleppegrell, 2008) with an increase in language proficiency or reading level, these data did not support that assumption. AGL students used an average of 2.11 clauses per T-unit, but the average for Below Level (BL) students was nearly the same, at 2.06. OL students averaged 1.95 clauses per T-unit. APP students had the highest average number of embedded clauses per T-unit, 0.57, followed by BL students at 0.51; AGL students averaged 0.45 on this measure. Meanwhile, in terms of ELPA groups, students at Levels 1 and 2 displayed slightly less complexity in their writing than did either those at Level 3 or the monitored students, using an average of 1.85 clauses per T-unit versus 2.04 and 2.03, respectively; average numbers of embedded clauses per T-unit were 0.32 for ELPA 1s and 2s versus 0.54 for ELPA 3s and 0.48 for the monitored group.²³

These differences should, of course, be interpreted cautiously given the size and composition of the sample, which includes subgroups consisting of as few as four students. Taken at face value, the data seem to show that, at least for this group, language proficiency and reading level did not predict performance on text complexity measures.

Going beyond numerical values reveals more about the ways students in the various subgroups tended to use embeddings. Some clause types were distributed fairly evenly across groups. These included the dependent clauses associated with conditionals,

²³ Total number of T-units for the ELPA Levels 1 and 2 groups was 15.80 versus 12.78 for Level 3 and 14.92 for the monitored status students. Total numbers of embedded clauses were 4.80 versus 6.33 for ELPA 3s and 7.33 for the monitored group.

as discussed in the previous section. They also included nominalized clauses, in which something already introduced into the text is presented in an embedded clause that functions as a nominal group, as in “[Using meat ants] is a good solution.” Though there were only a few instances of this clause type, the presence of nominalization is a sign that writing is becoming more advanced. This is part of the increased range of dependent clause types expected to emerge in the writing of students from ages nine to twelve (Christie & Derewianka, 2008). Other emergent clause types, such as those of purpose, reason, and manner, were noted too infrequently to draw meaningful conclusions about how their use may be influenced by language or reading level.

By contrast, the overwhelming majority of students across groups used embedded clauses in introducing their Evidence, e.g., “The first reason is [that toxins from the toads are not likely to affect the ants].” Students in the two higher reading groups, however, were more likely to include an additional clause connecting Evidence to Claim, as in “One reason [why scientists shouldn’t use the meat ants] is...”. The additional reiteration of Claim done by the higher-level students is an effective way of showing that Evidence has been appropriately selected. This move was never explicitly discussed, though it was present in the model arguments; the tendency for higher-level students to include it in their writing would seem to reflect a sensitivity to the purposes, or at least the shape, of the argument genre as modeled in the unit.²⁴

Students in the lower two reading and language proficiency groups, for their part, used more projected clauses. Projections frequently present ideas, in this case almost always following the phrase “I think”, as in “That is why [I think / we should study more].” The phrase “I think” was used very few times by the higher-level students. This

²⁴ In a pre-instruction argumentative writing task, this feature was present in only one student response.

would again seem to point to a higher sensitivity to the linguistic demands of scientific registers, which tend to value a depersonalized, authoritative authorial voice, characterized in part by the absence of first-person pronouns and the absence of making one's thinking process explicit. This point was also not discussed explicitly in the unit, and it is unclear whether this is an idea students had been exposed to elsewhere. Nonetheless, it is another example of the range of sophistication present in the writing of these fourth-grade students, and of the kind of difference made visible by thorough, descriptive linguistic analysis.

Trends: SFL metafunctions. The following section describes the fourth graders' use of ideational, interpersonal, and textual resources in composing their arguments.

Ideational resources. The argument genre presented in the unit is characterized by the use of generalized participants (i.e., "cane toads" as a class rather than naming a specific participant), a pattern reflected in student writing. Participant analysis showed that the reference chains²⁵ used by the students concerned generalized participants (cane toads, meat ants, poison/toxins, caterpillars, honeydew, carcasses, lungworms). The students' writing also conformed to expectations that the majority of processes used would be *being* and *doing* processes, used in this case to describe and define elements of the cane toad crisis and offer opinions on which actions should or should not be taken and why. Students used an average of 17.23 *doing* processes (0.59 per clause, or PC) and 7.27 *being* processes (0.26 PC) in their arguments, compared to 0.38 (0.01 PC) and 2.19 (0.09 PC) *saying* and *sensing* processes, respectively.

²⁵ Reference chains consisted of two or more uses of the same participant. Instances of "I", e.g., "I think" and "[the Australian] scientists", e.g., "Scientists should use meat ants" were discarded because their function was rhetorical; they did not contribute to the central ideas being developed in student writing.

Saying processes in argument typically do not report direct speech, but are instead used to recognize a counterargument or to treat the evidence as a semiotic *sayer*. Both of these uses were observed in the fourth-grade data. The phrase “Some people say...” accounted for just one of the ten *saying* processes used by fourth-grade writers. Most others used more complex formulations of the verb such as “You might say” or “Some people may say”; one student’s *saying* process came in the form “My evidence states...”

Sensing processes were used more frequently than *saying* processes, with the process “think” accounting for nearly 95% of the *sensing* processes employed. *Sensing* processes are used to report on internal thinking. This process type is common in the texts produced by young elementary schoolers, but over time students are encouraged not to report on themselves as thinkers in this way, especially in the context of scientific writing. Notably, while some students’ use of *sensing* processes concerned their own thoughts, the process type was also used in many cases to report on the beliefs and expectations of various stakeholders in the cane toad situation (“they *thought* cane toads would solve the problem”). As the clause-level analysis suggested, use of *sensing* processes decreased as language proficiency level increased, moving from 0.10 PC at Level 3 and 0.09 PC at Levels 1 & 2 to 0.06 PC for monitored students. This suggests there may be some minor development in this area accompanying an increase in language proficiency, though these numbers should be interpreted cautiously.

No patterns were identified regarding concentration of *being* vs. *doing* processes according to reading or language proficiency level. Students at all language proficiency levels used roughly the same concentration of *doing* processes. Monitored students used

an average of 0.03 PC fewer *being* processes than did students at Level 3 (0.25 vs. 0.28, respectively).

The range of ideational resources used to establish logical relationships was significantly expanded in the fourth-grade students' texts compared with texts from the second grade. Conjunctive resources associated with *contrast*, *condition*, *cause/reason*, *addition*, *time/sequence*, *purpose*, *concession*, and *consequence* were observed, with the first four meaning types accounting for nearly 75% of all usages. No trends related to reading level were identified, but the data showed some minor differences between language proficiency groups in the use of some meaning types.

Contrastive meanings, which made up 21.62% of all conjunctive meanings, included "but" and, to a lesser extent, "however":

You might think that meat ants protect caterpillars but when the meat ants go to look for
food the caterpillars might be eaten. (Sample 47)

Usually the poison in the frog kills its predators. However, the poison in frogs doesn't
affect ants. (Sample 16, Mon/AGL)

They brought 102 toads to solve a problem but they became an even bigger problem
themselves. (Sample 137, Mon/OL)

Use of contrastive meaning rose slightly with increases in language proficiency, from 0.04 PC at Level 3 to 0.08 PC at Levels 1 & 2 and 0.10 PC for monitored students. This is notable because, as seen in the examples above, the language of *contrast* played an important role in the presentation of facts and hypotheticals central to many students' arguments. Contrast, of course, is the essence of the Evaluation of Counterargument, but its use was by no means limited to that stage; it frequently appeared in Claims and

occasionally in the Description of Problem, as in Sample 137, where it subtly introduced the concept of unintended consequences. Overall, the widespread use of connectors of *contrast* indicates writing consistent with the goals of the genre.

Conjunctive “if”, as might be expected, constituted all of the *conditional* meanings noted in the sample. This “if”, associated with the subordinate clause in the conditionals emphasized in the argument unit, was used at roughly the same per-clause rate by all language proficiency groups. Sometimes connected to the conditional were meanings related to *cause/reason*, usually expressed using “because”, as in the following example:

I think if we take away the ants then the caterpillars might die because the meat ants are
the ones that protect the caterpillars. (Sample 94, 3/BL)

Use of connectors of *cause/reason* decreased slightly as English proficiency increased, from 0.08 PC at Level 3 to 0.07 at Levels 1 & 2 and 0.04 PC for monitored students. “Because” was generally used in the same ways across language proficiency levels, but it is possible that the difference in per-clause concentration could be attributed to some students’ choices to omit “because” while otherwise preserving the expression of a logical relationship (i.e., juxtaposing two logically connected statements without using the connector to signal the relationship). While explaining reasoning remains central to the task of constructing an argument, students at this level have an expanded range of linguistic resources upon which to draw in doing so.

Interpersonal resources. Analysis of interpersonal resources involved examining *grammatical mood* and use of *modality*.

Mood. The fourth-grade writing data included declarative, interrogative, and imperative clauses. Only the first would be considered typical of the argument genre, and indeed, almost all of the clauses were of this type. Just four students departed from this pattern, three using interrogatives and the other using an imperative. The atypical *mood* choices are intentional rhetorical flourishes:²⁶

What if they don't eat the cane toads? We don't know if they will eat the toads, so the humongous toad population could get larger. (Sample 116, Mon/OL)

I also think it is a good option and they should study more because what if the meat ants invaded Australia and took over that country... would it be the end of Australia and its animals? (Sample 151, Mon/OL)

But wait don't act now we should study more about meat ants. (Sample 61, Mon/OL)

The writer of Sample 151 leaves her rhetorical question unanswered, while the writer in Sample 116 poses a question and immediately goes on to answer it. While Sample 151 adopts an appropriate register throughout the essay, Samples 116 and especially 61 display more typically conversational language features. The question of whether the arguments are strengthened by their authors' *mood* choices is open to interpretation: while they display features related to "voice" typically valued in evaluation of elementary-level writing and for that reason may be more interesting to read, those same features could be seen as detracting from the goal of presenting oneself as a serious participant in a scientific debate.

²⁶ With the exception of one student who appears to have misunderstood the guiding questions on the Counterargument/Evaluation scaffold, incorporating them directly into his argument. For example, he wrote: "The part I agree in the counterargument is that ants fight other ants for their territory. How much do I agree? 5." (Sample 90, 3/APP)

Modality. Three types of modality were reflected in students' writing: modality of *obligation, ability, and likelihood.* Modality of obligation was used to make recommendations about how the cane toad problem should be handled. This type of modal meaning was typically found in the Claim and Restatement of Claim, though as discussed above, many students used the language of obligation to reiterate their positions throughout their arguments. 92.31% of the sample, or 24 students, used modality of obligation (e.g., "Scientists *should* use meat ants but they *should* study more"). The two students who did not use a truncated Claim:

That's why I think good study more. (Sample 47, 3/APP)

That's why I think meat ants are a good solution. (Sample 143, 3/APP)

The writer of Sample 143 never explicitly connected the meat ant "solution" to the cane toad problem, and her argument suffered as a result. Sample 47's author wrote an otherwise fairly good argument; the infelicitous-seeming "good study more" is in fact a reflection of the repeated use of this phrase in classroom discussion as verbal shorthand for "meat ants are a good solution that should be studied more." It could be said that this is a case of implied obligation, though of course were further editing to be done on this piece the writer would be encouraged to include the longer version of the Claim, in which "should" actually appears.

The language of *likelihood* (e.g., "the cane toads *might* die"), which had been repeatedly emphasized throughout the argument unit, was also used by a majority of students: 88.46%, or 23 students. This type of modality was typically found at the Evidence/Reason and Counterargument/ Evaluation stages, where it helped to communicate how the writer's recommendations were informed by the likelihood of

different possible outcomes. Three students did not use this type of modality in their arguments. One struggled with several elements of the argument and may have missed or misunderstood the conversations around likelihood. The other two writers were able to convey a sense of possibility using modality of *ability*:

Meat ants are not affected by the cane toad poison, and with one bite the cane toad can die in 24 hours. (Sample 1, Mon/OL)

It is true meat ants can fight with many other ants. But they do not always fight other organisms. (Sample 85, 3/BGL)

The language of *ability* was used by precisely half of the students in the sample. That this figure is lower than those regarding modality of *obligation* and *likelihood* is not surprising; while modality of *ability* can be useful in presenting information about cane toads and meat ants (“meat ants can kill cane toads because they are not affected by the toxins”) or discussing what scientists could do to test or fix the problem (“scientists could grab a toad and an ant and see the results”), this type of meaning is less central to argument-specific goals. Overall, the fourth-grade data show that students are using an expanded repertoire of modal resources to convey different kinds of meanings, though there is still a fair bit of variation with respect to how effectively these tools are used.

Textual resources. As previously discussed, *grammatical theme* is the point of departure for the clause. When the *theme* is also the subject of the clause, it is referred to as *unmarked*, while a *marked theme* foregrounds something other than the topic of the clause; marked themes are a hallmark of writing development (Halliday & Matthiessen, 2004). In keeping with the trend toward an expanded range of linguistic resources, the fourth-grade sample contained more variety in *theme* types. There were still only 14

instances of marked *themes* in the fourth-grade data. Four involved a circumstantial element in theme position, as with the following examples:

And with one bite the cane toad can die in 24 hours. (Sample 1, Mon/OL)

In 1930s, 102 cane toads were brought to the country to solve a problem. but now there are as many as 2,000,000,000. (Sample 17, 3/BGL)

The remaining marked themes were examples of *interpersonal themes*, offering the writer's comment or perspective on what is about to be said as the point of departure for the sentence:

From my point of view scientists should use the meat ants, now! (Sample 137, Mon/OL)

[A]nd maybe the toad population will increase and Australia could have a very very huge problem. (Sample 116, Mon/OL)

Long story short, yes I think meat ants are a good solution to kill the cane toads. (Sample 6, 1-2/OL)

In my opinion it is good but they should study. (Sample 58, 1-2/BGL)

There are no particular expectations in argument regarding *theme*. Both the increase in marked *themes* and the continued prevalence of simple *unmarked themes*, however, are consistent with expectations for writers at this developmental stage (Christie & Derewianka, 2008).

Summary. The linguistic choices made by these fourth grader writers were largely consistent with both the demands of the argument genre and the kind of writing that is typical for writers of this age. Ideational meaning was built through reference to generalized participants; the use of mostly *being* and *doing* processes, in keeping with the need to present factual information in support of an argument; and through connectors of

contrast, condition, and cause/reason, which function to present explanations of argumentative logic and discuss the possible consequences of the recommendations being discussed. Interpersonal resources related to grammatical *mood* and modality were also genre-appropriate: almost all clauses were declarative, and students used modality of *obligation* and *likelihood* to make recommendations and predictions. Textual resources, while not dictated by genre, were used in developmentally appropriate ways.

In the next section, the focus of analysis turns to how students constructed meaning at the *stage* level.

Stage Analysis

Genres are described in terms of their social purposes, stages, and language features typically used to build meaning. The preceding section showed that fourth-grade writers used language features that were characteristic of the argument genre. This section places those observations in a more meaningful context by moving from an examination of linguistic features in isolation to a stage-by-stage consideration of the ideas they were used to construct.

The argument genre presented in fourth grade included seven stages: Description of Problem, Claim, Evidence, Reason, Counterargument, Evaluation of Counterargument (Evaluation), and Restatement of Claim. The analysis below considers most of these stages in pairs: Description of Problem/Claim, Evidence/Reason and Counterargument/Evaluation. Restatement of Claim was evaluated as a single stage. These pairings reflect the way the stages were presented in classroom instruction and scaffolds, which were designed to reinforce the interdependencies between stages. While observations about individual stages within the pair can (and should) be made, taking the

stage pair as the main unit of analysis gives a fuller picture of the degree to which the writer met the social and rhetorical goals of each phase of the argument.

To evaluate the fourth-grade writing, transcribed arguments were first divided by stages. All instances of stage pairs were grouped and placed into separate documents (e.g., one document for Description of Problem and Claim, another for Evidence and Reason, etc.) These documents were then reviewed and stage pairs were labeled as Successful, Partially Successful, or Unsuccessful based on an evaluation of the extent to which the writer met expectations for each stage. The criteria used to make these assessments will be presented below as each stage is discussed. The designations reflected my evaluation of performance in absolute terms but also relative to the rest of the writing samples, and the labels should be understood in terms of what they reveal about additional support students may need. A “Partially Successful” label, in other words, may be applied to a stage pairing that, independent of the context of this analysis, would be considered an example of good fourth-grade writing. For the purposes of this study, a detailed evaluation of student performance stage by stage is called for to aid in developing an understanding of how fourth graders begin to construct the argument genre.

As with the second-grade texts, reliability was established by having a member of the research team use the coding manual (see Appendix M) to assess 20% (n=6) of the class’s transcribed and anonymous arguments. This process consisted of assigning Successful/Partially Successful/Unsuccessful labels to the Description of Problem/Claim, Evidence/Reason, and Counterargument/Evaluation stage pairs, and to the Restatement of Claim stage. The exact agreement rate between my scores and the second rater’s was

87.5% for the fourth grade texts, and as such coding for this analysis was deemed reliable. The three instances of disagreement between raters involved minor differences in our perceptions of the writing, resulting in adjacent ratings for two Evidence/Reason pairs and one Counterargument/Evaluation pair.

Description of Problem, Claim, and Restatement of Claim. The prompt that fourth graders responded to was: *Scientists in Australia think using meat ants might be one effective way to stop cane toads from spreading and hurting more organisms. What do you think of this option? Is it a good option Australia should act on now, a good option they should study more, or not a good option? Why?* After considering the evidence, most students opted for a middle-of-the-road position: 17 of 26 students said that meat ants were potentially a good solution, but one that needed to be studied further before unleashing them into the wild. Six thought the cane toad issue was pressing enough, and the existing evidence that meat ants would work convincing enough, to advocate for using the meat ants now. Just two said the meat ants shouldn't be used.

Classroom materials given to students defined Description of Problem as a stage that *"familiarizes the reader with important background on the issue or problem."* Claim was described as *"a statement that introduces the specific issue to be argued and the writer's position on that issue."* Most students did fairly well in writing their Description of Problem and Claim stages: 11 wrote Successful Description of Problem/Claim pairs, another 11 were Partially Successful, and just four were Unsuccessful in writing this set of stages. Table 5.3 shows the criteria used to evaluate writing for this stage pair. To the extent that the Description of Problem/Claim posed a challenge for the fourth-grade writers, it was primarily one of scope: how much or how little detail to include. At

minimum, a Successful response would have to establish that cane toads were a problem for Australia, explain why, and introduce meat ants as a possible solution. This could be accomplished effectively in as few as two or three sentences:

Cane toads are a huge problem in Australia. They've been killing a lot of animals with their poison. Scientists think using meat ants will solve the problem. Meat ants will solve this issue. (Sample 151, Mon/OL)

Sample 151 presents cane toads as the problem, states why they are a problem, establishes that meat ants are a potential solution, and makes a claim, thus giving the reader the information needed to form a basic understanding of the problem.

The source texts, of course, provided additional information that could be included in a Description of Problem, such as dates, facts about the beetle problem, numbers of toads initially and at present, and any number of details about the meat ants themselves. Adding in such details at this stage rather than saving them for later in the argument sometimes resulted in a slightly unwieldy description, as in the following example:

Some people don't know about cane toads. But they are really poisonous, touch you're dead. Cane toads were brought to Australia to get rid of the beetles but it backfired and the cane toads became a big problem. Scientists tried to find ways to get rid of them. They came up with two solutions. First was lungworms. If they put them in the toad's body they would eat their insides. The second solution is meat ants. The cane toads' poison can't affect the ants and they can eat up to two toads in their lives. I think that they should use the meat ant it's better. (Sample 111, 1-2/BL)

Sample 111 was unique in scope, contextualizing the cane toad problem as one that scientists have continually struggled to address. (Only one other student mentioned lungworms at all, and did not go into any detail.) While the writer did a good job of setting up the issue overall, he gave the explanation of why cane toads are such a problem less attention than he ideally would have. It is clear that the toads' poison is problematic, but the stage pair could be improved upon by adding detail about the effects of the large cane toad population on the Australian ecosystem. On the other hand, there is arguably an excess of detail about meat ants at this stage (especially since this writer went on to repeat one of the details as his Evidence). Ultimately, despite the minor imbalance in level of detail, this argument met all of the basic requirements for a Successful response and was marked as such.

<i>Description of Problem and Claim</i>	
Successful	Establishes that cane toads were a problem for Australia, explains why, introduces meat ants as a possible solution, and presents the writer's position.
Partially Successful	Issues vary, but generally one of the following applies: Provides a partial description of the issue but leaves out a key element; includes all the necessary elements but gives insufficient detail; adds details that were not strictly relevant and proved to be distracting; does not include Claim (but writer incorporates position elsewhere).
Unsuccessful	Omits a significant amount of information and leaves the reader unprepared to contextualize the subsequent elements of the arguments.

Table 5.3. Criteria used to evaluate Description of Problem and Claim

On the other end of the spectrum, the four Unsuccessful examples included far too little information for readers without prior knowledge of the situation:

Meat ants... I think it's a good option but should be studied more. (Sample 84, Mon/OL)

The cane toads have been a big problem to Australia. Cane toads are not like other frogs or toads since they don't like looking for small insects. So they eat anything that will fit in their mouth. (Sample 143, 3/APP)

Sample 143 makes no mention of the search for a solution to the cane toad problem, or any mention of meat ants at all, and makes no Claim whatsoever. Sample 84,

on the other hand, is almost all Claim. The writer provides no hint of what meat ants are an option for and fails to even mention cane toads. Characteristically of Unsuccessful responses, these stage pairs omit a significant amount of information and leave the reader unprepared to contextualize the subsequent elements of the arguments.

Partially Successful Descriptions of Problem/Claim showed more variation in terms of what needed to be improved, but this designation was generally applied to stage pairings that provided a partial description of the issue but left out a key element; included all the necessary elements but gave insufficient detail; or overshot the mark by adding details that were not strictly relevant and proved to be distracting. This category includes four students who failed to make an initial Claim, but incorporated their positions into one of the other stages of their arguments.

As discussed above, issues of determining scope and summarizing appropriately extend far beyond what was addressed within the context of the argument unit. Analysis of these texts, though, does point to one language issue that could strengthen writing at this stage if addressed in future iterations of the pedagogical approach: the distinction between presenting and presuming reference. Eggins (2004, p. 33) explains that “participants in a text may be either *presented* to us (introduced as ‘new’ to the text) or *presumed* (encoded in such a way that we need to retrieve their identity from somewhere.” The position of the Description of Problem relative to the other stages makes clear that it calls primarily for *presenting* reference: there is simply nowhere else in the text from which novel participants’ identities may be retrieved. Nor are we dealing in this case with subject matter that is retrievable from local or cultural contexts, as the meat ants issue would be unfamiliar to most readers in the United States (including the

“class across the hall” discussed during the unit as a possible target audience). When this issue came up, it was typically around the introduction of meat ants. The following sample addressed meat ants using *presuming* reference:

Scientists should use meat ants because in the 1930s they brought 102 cane toads to Australia. Beetles were eating their crops, they thought cane toads would solve the problem. But the cane toads made it worse. The meat ants are a good solution and we should use them now to kill most of the cane toads in Australia. Scientists think meat ants are a good solution to get rid of the cane toads. (Sample 6, 1-2/OL)

Sample 6 referenced using meat ants in its first sentence, without having established what they are or what their significance is. Were this an example of cataphoric reference, in which the referent is subsequently clarified, the immediate appearance of “meat ants” would not be a problem. In this sample, however, the reader never gets more definition than the vague “good solution.” Compare this to an example of good use of *presenting* reference:

The cane toad has been a big problem to Australia. They were brought to the country to stop the beetles [from eating]²⁷ sugar cane crops. But they didn’t eat the beetles they ate anything that fits in their mouth. Scientists came up with an idea. They want to use meat ants. Meat ants are ants that eat meat. Meat ants are able to kill poisonous cane toads. The cane toads toxins are not likely to affect the meat ants. Also the cane toad mom can lay up to 35,000 eggs each year. (Sample 51, 3/APP)

The writer of Sample 51 organized his introduction of meat ants quite differently. Short, clear sentences with easy-to-follow referents are used to provide the reader with

²⁷ Bracketed text absent in original text; supplied here for readability.

useful background information. The contrast between how each of these writers brings meat ants into the discussion provides a useful illustration of how reference can contribute to the overall strength or weakness of the Description of Problem stage, and could serve as the basis for a mini-lesson drawing students' attention to and giving them practice with this language tool.

Finally, student texts at the Restatement of Claim stage displayed little variation. The description of the stage provided on instructional materials was “*Summarizes and restates the writer’s position on the issue.*” Without much instructional attention paid to this stage, the majority of students simply gave one-sentence restatements of their Claim (“That’s why I think meat ants are a good solution”, “This is why Australia shouldn’t use the meat ants to help get rid of the cane toad,” etc.). The 19 instances of this approach were marked as Partially Successful. Five students included their claim and an additional sentence reiterating or summarizing a key piece of their argument; these instances were marked as Successful. One student’s attempt consisted of repeating some facts about cane toads, departing from expectations to such a degree that it was marked as Unsuccessful, and one student omitted this stage entirely. Table 5.4 summarizes the evaluative criteria for this stage.

<i>Restatement of Claim</i>	
Successful	Summarizes and restates the writer’s position on the issue.
Partially Successful	Restates the writer’s position on the issue without additional summary.
Unsuccessful	Includes a final statement but neither summarizes nor restates the claim.

Table 5.4. Criteria used to evaluate Restatement of Claim

The link between Claim and Restatement of Claim seems like a direct one, but surprisingly, only 16 of the 25 students who included the Restatement of Claim were consistent with their claims from initial to final stages. Three students amended their

claims from start to finish. Two who initially said meat ants were a good solution to the cane toad problem but needed to be studied more took a stronger stance in concluding their arguments (one wrote that cane toads should not be used, and the other that they should be used right away). The other student initially took the position that meat ants were a good solution, but noted only in concluding that further study was needed. Two students did not have a discernible claim at this stage, one of whom used the pronoun *this* to reference the entire argument (“If scientists or the government of Australia could take *this advice* into consideration...”). Of the four students who did not make an initial claim, two concluded that meat ants were a good solution and one wrote that meat ants needed to be studied more, while the remaining student failed to include a claim either initially or in concluding, instead incorporating her position into her Evidence/Reason. These variations were unexpected and indicate a need to explicitly address the relationship between the two stages in question, perhaps with the aid of a scaffold prompting students to be sure that they match.

Evidence and Reason. 23 students wrote arguments that included the Evidence/Reason stage pair, of which six were marked Successful, seven were Partially Successful, and ten were Unsuccessful. Nine students included two sets of Evidence and Reason. In these cases each pair was analyzed on its own merits, but students received just one overall marking. Table 5.5 presents the criteria used to label writing at these stages. Classroom materials described Evidence as “*a piece of information from the text that supports the overall claim*” and Reason as “*a statement about why/how the evidence supports the claim and interprets the evidence in relation to the larger issue.*” Students

who wrote Successful Evidence/Reason pairs fully complied with these expectations, as in the following example:

One reason why scientists shouldn't use the meat ants is that even though they're called meat ants they mostly eat honeydew from caterpillars and butterflies. They also scavenge for carcasses. Why the meat ants aren't a good solution is that they're scavengers, which means they eat dead animals and are not used to hunting live prey. It may affect the ecosystem and organisms living there. (Sample 138, Mon/AGL)

The writer of Sample 138, arguing that meat ants were a bad solution, began by providing facts about the meat ants' typical diet. She then went on to explain why she felt that made them a bad solution to the cane toad problem, referring to two potential problems mentioned in source materials: first, that scientists were unsure of the extent to which scavengers would take to eating primarily live prey, and second, that changing meat ants' diet and, therefore, their behavior would affect the ecosystem in unknown ways. The move to define the scientific term "scavenger"—a definition not provided in the source material—echoed the writing in informational texts found in elementary schools, indicating an awareness that a scientific register is an important element of this kind of argument. The last sentence does leave the reader wanting more precise information about the effects it alludes to. The presence of this weak spot in the argument is characteristic of Successful Evidence/Reason pairs; while students met expectations for these stages and crafted strong responses overall, there was frequently at least one point that would leave room for improvement were subsequent drafts to be composed. This problem is not unrelated to the questions of scope raised in the Description of Problem stage, and indeed, though the issue is most central to the success of that initial stage,

stage analysis highlights the extent to which the ability to determine what information is relevant and how much to include remains important throughout the argument.

<i>Evidence and Reason</i>	
Successful	Provides a piece of information from the text that supports the overall claim and a statement about why/how the evidence supports the claim and interprets the evidence in relation to the larger issue.
Partially Successful	A connection between evidence and claim is present but incomplete/ insufficient, or material following evidence does not function to make such a connection (e.g., reiterates claim or adds facts without interpreting).
Unsuccessful	Lacks reference to Claim or anything that could otherwise be interpreted as a Reason, or adds commentary that does not make sense.

Table 5.5. Criteria used to evaluate Evidence and Reason

In contrast to the relatively minor flaws exhibited by the Successful examples, Partially Successful Evidence/Reason pairs tended to have weaknesses that compromised the overall effectiveness. The following sample illustrates how such weaknesses were frequently manifested in Partially Successful writing:

One reason we should study this more before we use it is meat ants mostly eat honeydew from caterpillars. They also scavenge for carcasses on the ground. We should study this more because meat ants mostly eat carcasses and honeydew. Another reason we should study this more is cane toads rely on their toxins to kill their attacker. But the toxins are not likely to harm the meat ant. When an attacker attacks the toad, it just sits there and lets its poison harm the organisms. When the meat ant attacks, it can eat it easily because it is just sitting still. (Sample 26, Mon/AGL)

The writer of Sample 26 argued that meat ants were a good solution that needed to be studied more. Her first set displays reasoning akin to the incomplete bridging discussed in Chapter 4: in her Evidence, she established that meat ants' diet typically consists of honeydew and carcasses, and reiterated this in her Reason as if its relevance to the issue were self-evident. Where Sample 138 suggested that changing this diet could be

bad for the ecosystem but failed to take the extra step of explaining how, in Sample 26 the basic question of why the Evidence presented means that further study is warranted goes completely unanswered.

Likewise, what would be the second Evidence/Reason pair appears to be missing the Reason altogether. The lead-in to the Evidence (“Another reason why we should study this more is...”) sets the reader up to expect information about either an unknown or something with potentially negative consequences, but the writer goes on to share facts about how meat ants overcome the cane toads’ normally very effective defense system. In the absence of the introductory clause, one might assume that this was an example following Mr. J.’s suggestion of introducing one piece of Evidence showing that meat ants were good and another showing that they should be studied more, but without a Reason explicitly establishing the logic undergirding the inclusion of these facts it is impossible to know whether this was the intent. At any rate, while this student was articulate and seemed to have a good idea of the shape these stages were supposed to take, including the expected language features, she struggled to successfully establish the logical connections between key parts of her argument. These weak or missing links between Evidence and Reason or Reason and Claim were the phenomena that led to the identification of Partially Successful examples together at this stage.

Unsuccessful Evidence/Reason pairs in fourth grade exhibited one of a number of serious flaws. Six lacked any reference to Claim or anything that could otherwise be interpreted as a Reason, consisting solely of the writer’s reporting of facts (or, in one case, several sentences imported directly from one of the source texts). The remaining four made some attempt to comment on or interpret their Evidence, but were either not

able to do so in a way that advanced their argument or wrote something that simply did not make sense.

Overall, Evidence and Reason were challenging in different ways in the fourth grade. Although there was more evidence to choose from, most students gravitated toward one or both of the two pieces of evidence laid out in Sample 26, perhaps in part because these had been featured in classroom discussion. Whereas the Evidence stage in second grade consisted of a straightforward reiteration of a single, simple fact, fourth graders writing this stage were tasked with something more complicated. Not only did they have to assess a larger amount of source material and make judgments about the appropriate amount of factual detail to include, but their reporting of facts needed to take into account not a single characteristic but explanations of systems (in the case of the interdependencies between meat ants, their usual diets, and other organisms in the ecosystem) and/or processes (when describing the mechanism through which cane toads evade attacks by almost all predators except meat ants). Though a basic explanation was in reach for most students, without additional support the step of presenting the relevant facts alone was clearly difficult for even the highest-performing students.

The Reason stage, for its part, became more challenging with the introduction of the third, qualified position, “good, study more”, which required that students demonstrate both positive and negative aspects of meat ants as a solution to the cane toad problem. Tellingly, three of the six students who wrote Successful Evidence/Reason pairs were arguing either fully in favor or fully against using meat ants. For many other students, Reason was either missing or insufficient. Even students who chose the “good, study more” option seemed to be more comfortable supporting one of its two component

claims, generally the “meat ants are a good solution” piece. With the understanding that Reason tends to be a less intuitive element of argument, often absent in the oral arguments of children of this age (Anderson et al., 1997), it may be that these fourth graders—who, it bears repeating, were learning about and writing this kind of argument for the first time—would have benefitted from practice writing Evidence and Reason in support of a simple Claim before taking on the more sophisticated measured position. Given the second grader’s success with this stage, there is no reason to believe that it is beyond fourth graders’ reach; rather, the difficulties reflected in the data suggest that this is an area in need of further support.

Counterargument and Evaluation. 20 of 26 students included a Counterargument/Evaluation pair. Of these 20 students, four had Counterargument/Evaluation pairs that were marked Successful, five were Partially Successful, and 11 were Unsuccessful. As with the second graders, performance on this set of stages was not determined by how well students had achieved the goals of the Evidence/Reason pair. Of the four students whose writing was marked Successful at this stage, only one also had a Successful Evidence/Reason pair. Two had been Partially Successful and another had been Unsuccessful. This inconsistency in scores from one stage pair to the next was expected based on second-grade data as well as the results of a preliminary study examining fifth graders’ arguments produced in an FG-based instructional context (O’Hallaron, 2012). Even without these other findings, it would be reasonable to expect variation in performance from stage to stage given the functional differences between them.

Table 5.6 presents the criteria used to evaluate Counterargument and Evaluation. Counterargument was defined in class materials as “*a piece of information that someone presenting another side of the argument might use for evidence.*” Evaluation of Counterargument, according to the same materials, “*evaluates the strength of the counterargument and explains what this information means for the writer’s position on the issue.*” The scaffold for this stage (see Appendix D) prompted writers to identify a Counterargument; think about which part(s) of the Counterargument they agreed with and rate the strength of their agreement on a scale from 1-10; think about which part(s) of the Counterargument they disagreed with and rate the strength of their disagreement on the same scale; and write an Evaluation based on their weighing of the Counterargument in the intermediate steps. Not reflected on the scaffold, but discussed in class, is the fact that presenting a counterargument in full involves sharing both evidence that could be used to support a different position *and* a reason explaining why—meaning that the Counterargument/Evaluation stage pair involves the same skills required in Evidence/Reason applied to an opposing position, plus the ability to give a balanced explanation of why the reasonable counterargument is ultimately not convincing. That the group’s performance was lowest for this pair of stages, then, is a result that should be interpreted with this higher level of difficulty in mind.

<i>Counterargument and Evaluation</i>	
Successful	Provides a piece of information that someone presenting another side of the argument might use for evidence and evaluates the strength of the counterargument, explaining why the Counterargument does not change the writer’s position.
Partially Successful	Presents Counterargument and attempts an Evaluation but does not adequately explain why the Counterargument does not change the writer’s position (e.g., explanation is incomplete/logically weak, or concession to opposing position is excessive and does not support original position).
Unsuccessful	Presents Counterargument but either does not include anything resembling an Evaluation or adds commentary that does not further the argument (e.g., reiterates evidence or writer’s position on the issue, makes an unrelated statement).

Table 5.6. Criteria used to evaluate Counterargument and Evaluation

Most fourth-grade students were able to identify and present evidence for an opposing position. Successful and the stronger Partially Successful examples also tended to include a reason in their Counterargument. Most students used a *saying* (“Some people might say...”) or *sensing* (“You may think...”) process in initiating the Counterargument stage to signal that they were presenting someone’s view other than their own, and the failure to do so was associated with responses that were weaker in general: this language was missing in two of the four Partially Successful and five of 11 Unsuccessful examples. As with the second-grade writing, however, it was the strength of the Evaluation that ultimately determined the strength of the stage pair. Where the success of second-grade responses depended on selecting the single Counterargument that could be directly rebutted, fourth-grade writers adopted six distinct strategies in constructing their Evaluations. These strategies included direct rejection of the counterargument using facts; direct rejection of the counterargument using a hypothetical; direct rejection based on opinion; providing an alternate solution to address the concern raised in the counterargument; total concession of the point raised in the counterargument; and making an unrelated statement. As one might suspect, some of these strategies were more helpful than others.

Only two strategies were used by the four writers of Successful Counterargument/Evaluation pairs. The following stage pair provides an example of a direct rejection of counterargument using facts, a strategy employed by two of the four Successful writers and on eight occasions total throughout the sample. It is shown in italics here for emphasis:

Some people may say that we should not use the meat ant because they could fight with other ants. Also they can change the behavior of other organisms. It is true that meat ants can start to fight with other ants for their territory because meat ants are the dominant type of ant. *But they do not always fight.* It is also true that they can change the behavior of other organisms. *But it is not for sure that it is going to completely change the behavior.* (Sample 26, Mon/AGL)

The writer of Sample 26, whose two Evidence/Reason pairs were discussed above, again chose to address two points in her Counterargument/Evaluation. Her position was that meat ants were a good option that needed to be studied more, and accordingly, she reported on an objection that could be raised by someone arguing that the ants were a bad solution that should not be used. The first part of the Counterargument had to do with the observation that meat ants, a dominant species, could become aggressive and attack other ant species when relocated to new areas to address the cane toad problem. Following the scaffold and argument models, the student writer first made a partial concession to the opposing side, acknowledging that fighting between ant species could occur. She then moved to explain why this counterargument did not convince her, addressing it directly and rejecting it on a factual basis. Her (correct) reading of the language used to present this point in the source material told her that interspecies aggression could be, but was not always, a problem, and her presentation of these facts provided logical support for the claim that it would be wise to study meat ants further before using them. She adopted the same strategy to address the second piece of her Counterargument. This second pair was weaker, but this was a reflection of the similarly vague information provided in the source material. The Evaluation again turned

on interpretation of modality used in the original text (“This could change the behavior of other organisms”). Though more information would have been helpful in this case, the writer accurately presented the available facts in support of her claim.

The second most effective Evaluation strategy, providing an alternate solution to address the concern raised in the counterargument, was used in three of the four Successful Counterargument/Evaluation pairs, and was used only at this level. The following stage pair gives an example of this strategy, again italicized for emphasis: If someone might argue with me that said it is bad, don’t [use] it, they might say the meat ants eat honeydew from caterpillars and butterflies. So it will be hard to make them eat the cane toads. I agree that they eat honeydew, it is their main food. But I disagree that it will be hard to make them eat cane toads. *They can put honeydew on the cane toads.*

(Sample 58, 1-2/BL)

The writer of Sample 58, arguing the “good, study more” position, effectively signaled the presentation of his Counterargument and included both evidence for the “bad, don’t use” position (what meat ants’ normal diet consists of) and a reason tying the evidence to the opposing claim (cane toads are not part of the usual diet, and meat ants may not want to make the switch to toads as a primary food source). He then made a partial concession, agreeing with the fact used as evidence, but went on to immediately establish that he disagreed with the reason presented in the Counterargument. Here, instead of rejecting the counterargument on a factual basis, he explained why it was not convincing by offering an alternative solution. The solution, of course, presents its own practical concerns, but in the writer’s view it is an ostensibly plausible workaround to the challenge of imposing a drastic dietary change on the ants. Though the alternate solution

strategy is somewhat less elegant than the purely factual rebuttal, the writers who employed it demonstrated solid understanding of the source material, proposing solutions that—even if difficult to implement—took into account various factors outlined in the texts.

A similar strategy, in which speculation about a hypothetical consequence is used to reject the counterargument, was used once at the Partially Successful level and twice in the Unsuccessful group:

Some people might say meat ants can kill toads. The toxins aren't likely to hurt the ants. The toads just sit there. So the toad population could decrease. *What if they don't eat the cane toads? We don't know if they will eat the toads, so the humongous toad population could get larger.* (Sample 116, Mon/OL)

This writer argued that meat ants were a bad solution and should not be used. In response to the counterargument that the ants' immunity to cane toad toxins could make them highly effective in reducing the toad population, he poses a hypothetical question: what if it doesn't work? This example was marked as Partially Successful. The writer signaled the presentation of his Counterargument and included both evidence and a reason. He had a good idea of what was being asked of him in the Evaluation, wrote something that departed somewhat from the suggested format but made sense overall, and paid attention to the language of likelihood in crafting his answer. What made this example only Partially Successful was the Evaluation, italicized above, in which the writer suggested that meat ants ultimately might not work as a solution but failed to establish why this means scientists should not attempt to use them. The question he posed was valid, and the hypothetical is a potentially useful conceit. This strategy could in

theory be combined with rejection based on facts to build a convincing Evaluation, but the students who used it in this class did not do so. This use of a hypothetical to reject the counterargument instead essentially amounted to describing a scenario in which the counterargument would be rendered invalid. Continued growth of the cane toad population would result from the failure of *any* solution that was unsuccessfully implemented, and failure in this context is always a possibility, making the reasoning too general; this student's response to the counterargument, as it was written, was not specific to meat ants. Had Sample 116 provided additional detail pointing to *why* the success of meat ants isn't guaranteed—for example, by including the information in Sample 58 about meat ants' diet—it would likely have been marked as Successful. The writer was correct in arguing that we don't know, and correct in predicting growth in the cane toad population if meat ants don't eat them, but in the absence of stronger support for this hypothetical failure the reasoning in the Evaluation is incomplete.

Two Partially Successful students weakened their arguments significantly by crafting their Evaluations using a strategy of total concession, as in the following example:

Some people may say that the ants mostly eat honeydew from caterpillars. They scavenge carcasses. *This changed my mind because if caterpillars eat honeydew mostly, they may never eat cane toads again. They scavenge carcasses all the time so cane toads are probably the only living things the meat ants eat.* (Sample 16, Mon/AGL)

The writer of Sample 16 took the position that meat ants were a good solution that needed to be studied more, and in the course of her Counterargument and Evaluation appeared to have convinced herself otherwise. Though other attributes of the stage pair

indicate that the writer had a basic understanding of what she was expected to accomplish, when it came time to demonstrate that she had considered the issue she overshot the mark. If she had truly changed her mind, of course, she would need to make a different Claim and alter the rest of her argument accordingly. Assuming the Counterargument did not totally convince her to change her position on the issue, what is missing here is the explanation of why not.

Finally, two remaining strategies were only associated with Unsuccessful Counterargument/Evaluation pairs. The use of opinion to reject the counterargument was observed just once:

Some people say it's worth trying *but in my opinion the scientists should study it more.*

(Sample 84, Mon/OL)

Though the writer of Sample 84 presented something recognizable as an opposing view and his response to it, this approach to writing the Evaluation was essentially devoid of substance, the functional equivalent of a weak Restatement of Claim. Slightly more substantial, but still ineffective, were the three cases of Evaluations that were entirely disconnected from their Counterarguments:

The meat ants normally eat honeydew from caterpillars and butterflies. The meat ants might eat the cane toads sometimes but normally they eat the honeydew. I strongly disagree because most of the meat ants' bites kill the cane toads immediately. (Sample

13, Mon/AGL)

The writer of Sample 13, arguing that meat ants should be used now, presented his Counterargument without indicating that it was somebody else's concern. He then responded to the point that meat ants normally eat honeydew by pointing out,

inaccurately, that meat ants' bites kill the cane toads immediately. The lack of connection between the two stages prevented them from building on one another in the way that was intended, rendering them ineffective in combination with one another.

It was somewhat surprising to see the range of strategies used to formulate Evaluations in the fourth grade, given that model texts and classroom discussion emphasized the factual approach. Some of the strategies would be best treated as misunderstandings of the genre goals or the tools provided to help with writing; others seem to have arisen as an unintended result of classroom discussion. A detailed discussion of the fourth graders' strategies for writing the Evaluation and implications for pedagogy will be presented in Chapter 6.

Framework for Evaluation of Argument in Fourth Grade

The expanded features of the fourth-grade curriculum and writing necessitated some modifications of the evaluative framework and procedure presented in Chapter 4. This section discusses the differences between the second- and fourth-grade frameworks, which arose following analysis showing what students at different performance levels actually do in writing each stage pair in the two grades. Using this analysis as the basis for the evaluative framework allows teachers to pose key questions about the quality of the structure, logic, and language used at each stage of student's argument.

The framework is presented in Table 5.7, below. The five dimensions comprising this framework are: Logic, Factual Correctness, Consistency, Function, and Features. Logic and Factual Correctness are straightforward and not particular to an SFL-based approach. They ask, simply, whether the stages being evaluated make sense and accurately reflect source material and/or classroom discussion. Function refers to whether

the stage is generally recognizable—whether what is written functions to move the piece forward in a way that aligns with expectations. While descriptions for these first dimensions are the same throughout the written argument, specific descriptions of Consistency and Features vary by stage pairing.

Evaluative Dimension	Criteria: Evidence and Reason	Criteria: Counterargument and Evaluation
I. Logic	Does it make sense? Is the reader able to follow the argument without needing additional background?	Does it make sense? Is the reader able to follow the argument without needing additional background?
II. Factual Correctness	Do(es) the statement(s) accurately reflect the source material and/or classroom conversation? Does the writer understand (is the writer able to explain) the concepts?	Do(es) the statement(s) accurately reflect the source material and/or classroom conversation? Does the writer understand (is the writer able to explain) the concepts?
III. Function	Is the general shape of the stages discernible? Do Evidence and Reason “do” what they’re supposed to? Does the author use language to signal movement between stages?	Is the general shape of the stages discernible? Do Counterargument and Evaluation “do” what they’re supposed to? Does the author use language to signal movement between stages?
IV. Consistency	Do Evidence and Reason, in fact, support the writer’s claim? Does the Reason address the particular piece of evidence presented?	Is the Counterargument one that would be plausible for someone with a different claim to make? Is it, in fact, counter to the argument the writer is making? Does the Evaluation effectively address the counterargument presented (do the stages “match”)?
V. Features	Does the writer use language features associated with these stages where appropriate (e.g., modality to make predictions, discuss the likelihood of a particular outcome or how certain scientists are about facts)?	Does the writer use language features associated with these stages where appropriate (e.g., modality to discuss the likelihood of a particular outcome or how certain scientists are about facts, use of contrastive or concessive resources as needed)?

Table 5.7. Evaluative framework for argument (fourth-grade version)

Consistency in Evidence/Reason is about the internal coordination of ideas: whether Evidence and Reason actually support the Claim. In the second grade this was a relatively straightforward proposition, closely related to performance on the Function dimension, but in the fourth-grade writing it became clear that there is considerably more variation in how students connected Evidence and Reason to Claim. Given that over half the class chose a Claim that occasioned a more challenging, qualified response (one

explaining why meat ants are a good solution, but not good enough), it was perfectly possible for students to construct an Evidence and Reason pairing that satisfied all other expectations but was not crafted in a way that offered support for the stated Claim.

Consistency for Counterargument and Evaluation also became more complex in fourth grade, asking not only whether the Counterargument actually represents an opposing position and whether the Evaluation addresses the issue raised in the Counterargument, but also examining whether the Counterargument represents a concern that someone on another side of the argument would realistically have. With more potential source material to draw on, credible writers of arguments have to weigh potential Counterarguments and decide which is truly worth engaging with. The concern here is not so much with the creation of straw men, but with *relevance*. Students were asked to think of the strongest support for the opposing position and to use that in their Counterarguments. This part of the Consistency dimension asks if they have done so.

Finally, Features is a finer-grained examination of the students' use of relevant language features, with an eye toward appropriateness. Though appropriate or effective use may vary depending on stage, the language feature most heavily emphasized in instruction was the use of modality to express likelihood, comment on certainty, or make predictions (generally through the use of a conditional). The importance of this dimension derives not from a surface-level determination of whether writing is grammatically correct, but from the function the language features serve in realizing the argument. Poor marks in the Features dimension tells the evaluator that something is amiss, but further attention to the matter is necessary to determine whether the problem is one of language, logic, or lack of familiarity with the genre.

I used this framework to evaluate the students' text. First I classified argument as Successful, Partially Successful, or Unsuccessful, as discussed in the previous section. To provide further detail about the strengths and weaknesses of the writing samples, each dimension/question was marked as (Y)es, (S)omewhat, or (N)o. For arguments that included multiple pieces of Evidence and Reason, each pair was evaluated separately, yielding two Y/S/N marks for each dimension. As in the second grade, the purpose of the evaluative framework for argument was to assess the strength of the argument across multiple levels of meaning, not to be a rubric in the traditional sense. Marks in response to the questions were not meant to be quantitative "scores", but they did provide an interesting picture of student performance, one that tended to support the broader groupings.

Comparing framework scores to the performance levels in the initial stage analysis supports the Successful/Partially Successful/Unsuccessful classifications and underscores the potential utility of this more detailed instrument in classroom and research settings. While there were not hard boundaries between groups, the majority of marks on individual dimensions were Ys for the Successful group and Ns for the Unsuccessful group. Stages that had been marked as Partially Successful were characterized by a mix of Ys, Ss, and Ns, the balance of which shifted from student to student. Just as very few students had the same Successful/Partially Successful/Unsuccessful classification on all stages, very few students received all Ys, Ns, or Ss within a stage. As is the case with most kinds of writing, all students would have been able to improve their work somehow. This highlights a key contribution of the framework: where traditional assessments guide raters toward umbrella scores that may

over- or underestimate student performance on specific aspects of argumentation, and provide only vague support for helping students improve their work, the evaluative framework is designed to pinpoint specific strengths and weaknesses across and within stages. In classroom application, the individualized distribution of Y/S/N markings would help teachers provide targeted feedback by identifying exactly what a student has done well and calling attention to the specific aspects of a particular stage that were in need of additional revision.

The following section introduces three focal students and shows how this method of evaluation relates to their writing, illustrating how the framework could help teachers go beyond generic rubrics in providing feedback on students' work.

Focal Students

Isa. Isa was a child with bright eyes, a bright smile, and a seemingly inexhaustible supply of energy. He participated actively in class discussions, though in small group or individual work his attention quickly wandered. He read on grade level. From kindergarten to third grade he had tested at the highest level of proficiency within the ELL program, and with the Spring 2012 administration of the ELPA he officially exited to monitored status. Like other focal students, Isa had been identified by Ms. Haddad as a student who was engaged in general and “really, really into Functional Grammar.” Isa was the student who, on Day 2 of the argument unit, had pointed out to Mr. J. and the rest of the class that what we called Counterargument carried out a very similar function to Evidence and Reason. He also had an extended interaction with Mr. J. on Day 4 (see description of instruction above) in which he gave an articulate and creative explanation as to why a piece of evidence the class was discussing supported a particular position.

Given the understanding he displayed during the argument lessons, it is not surprising that he went on to write one of the strongest pieces in the class. Isa's argument showed a strong sense of awareness that argument is a highly interactive genre, and this awareness translated to his responses in an interview conducted at the conclusion of the argument unit. He was easily able to identify the different stages in his own piece and was conversant in terms of his purposes for writing and the science behind it. He was also able to explain his choice of "likelihood words", but keyed in on phrases rather than the words themselves. Overall, Isa's responses in the interview were among the strongest in the focal student group as regards explicit knowledge of the genre structure.

Isa's argument, not unlike Isa himself, had quite a bit of "personality". The original argument is shown in Figure 5.1 below. The transcribed version, with spelling and punctuation standardized, reads as follows:

In the 1930s we all knew that beetles were eating sugar cane. Then we thought cane toads would solve the problem but they were a problem themselves. Now we have a new solution, it's meat ants. But wait don't act now, we should study more about meat ants.

Meat ants are the only organisms that will not be harmed by toxins. This proves meat ants are a good solution because if they're not harmed by the toxins, the cane toads would eventually die, but the meat ants would be fine.

What you might say is that meat ants protect caterpillars and if they don't protect them the caterpillars will be eaten and die. I understand that you disagree with me but we should study them more because we could take half of the meat ants population and put them near cane toads and leave the rest where they are to protect the caterpillars.

This is why I personally think meat ants are a good solution to be studied.

In stage analysis, Isa's Description of Problem/Claim pair was rated as Successful. He established that cane toads were a problem, gave some background information on how the problem came about, presented meat ants as a potential solution, and shared his position. The repeated use of "we" at this stage, along with the imperative in the Claim, are unique, functioning to quickly erase the anticipated distance between writer and reader. As previously discussed, a more sophisticated writer would be expected to adopt a more authoritative, distant stance, but few of the fourth-grade writers did so. Were this stage pair to be revised, additional information about the problems posed by the cane toads would be desirable, but overall the goals for these stages were achieved. His Restatement of Claim was rated as Partially Successful, consisting of a simple restatement of his position without summarizing additional content.

Isa's Evidence/Reason pair, shown in the second paragraph above, was one of the strongest Partially Successful examples. He chose a piece of evidence in support of meat ants being a good solution but did not include another suggesting they need to be studied more, a pattern typical of other fourth graders who had chosen the "good, study more" option. He clearly laid out how meat ants' resistance to toxins would make them good candidates to combat the cane toads, choosing language to convey that he felt certain about what he was saying. (His interview confirmed that this was an intentional choice.) The flaw weakening the effectiveness of this Evidence/Reason pair is insufficient information about cane toads' toxins. For a reader unfamiliar with this issue, the use of presuming reference in the Evidence ("toxins") is disorienting. Neither before or after this point was it clearly established that cane toads produce potent toxins that can kill

almost any predator, a fact that would be crucial to fully understanding the logic in this stage pair.

I In the 1930s we've all knew that beetles were eating sugar cane then we thought cane toads would solve the problem but they were a problem themselves.

Now we have a new solution it's meat ants but wait don't act now we should study more about meat ants.

Meat ants are the only organisms that will not be harmed by toxins

This proves meat ants are a good solution because if they're not harmed by the toxins, the cane toads would eventually die, but the meats would be fine.

What you might say is that meat ants protect caterpillars and if they don't protect them the caterpillars will be eaten and die. I understand that you disagree with me but we should study them more because we could take

half of the meat ants population and put them near cane toads and leave the rest were they are to protect the caterpillars.

This is why I personally think meat ants are a good solution to be studied.

Figure 5.1. Isa's handwritten argument

When the evaluative framework was used to apply Y(es)/S(omewhat)/N(o) labels for each dimension, this stage pair received Ys for Features and Function: Isa's Evidence

and Reason did have the overall shape and function they were expected to, and he used modality to indicate his certainty about the outcome he predicted. The other three dimensions—Logic, Factual Correctness, and Consistency, received Ss. The S in Consistency reflects the use of a single Evidence/Reason pair showing that meat ants are a good solution and the absence of another demonstrating the “study more” aspect of the claim, although in this case it is a qualified S, as will be discussed momentarily. Finally, the Ss in Logic and Factual Correctness point to what is most in need of improvement. The stage pair did make sense overall and it was not factually inaccurate, but some details were not fully explained and additional background information is needed.

Isa’s Counterargument/Evaluation pair, shown in the third paragraph above, was classified as Successful. His Counterargument was one of the few in the class that included both evidence (“meat ants protect caterpillars”) and reason (“if they don’t protect them the caterpillars will be eaten and die”). He constructed this stage pair as a direct interaction, signaling the presentation of his Counterargument by addressing a “you” with whom he is arguing. This interaction continues with his Evaluation:

I understand that you disagree with me but we should study them more because we could take half of the meat ants population and put them near cane toads and leave the rest where they are to protect the caterpillars.

This Evaluation is an example of the use of an alternate solution to reject the counterargument. Isa acknowledged his opponent’s position, and then went on to offer a solution that would satisfy both parties: conducting an additional scientific experiment in which some meat ants would be introduced into the cane toads’ environment while others would stay where they were. This suggestion was not hinted at in the source material but

actually seems like a sensible step to take towards minimizing the negative effects of ants' behavioral changes on other members of the ecosystem. Interestingly, Isa inserted a clause ("but we should study them more") into his Evaluation that would seem to address the issue raised at the Evidence/Reason stage regarding insufficient support for his claim. This was an innovation on Isa's part that turned out to be effective: he provided the Evidence and Reason showing that meat ants were a good solution, and let his imagined opponent set him up to explain why further study might be needed.

Isa's Counterargument/Evaluation received Ys on all dimensions of the evaluative framework (in addition to addressing the S assigned for Consistency in the Evidence/Reason pair). The writing in this stage pair was logical, factually sound, and functionally effective; the Counterargument was plausible and was effectively addressed through the Evaluation; and attention was paid to appropriate use of language features. Isa clearly signaled transitions between stages and was intentional in his use of modality to convey likelihood ("will be eaten and die"), obligation ("we should study them more"), and ability ("we could take half the meat ants...").

Isa's argument paints a picture of the kind of creativity and innovation that can emerge, even in a highly-scaffolded instructional context, when a student has a solid understanding of both genre-specific expectations and scientific content. Of course, this was not the case for all students. In the following sections are discussions of the ways in which different kinds of misunderstandings made this task challenging for two other focal students.

Sanaa. Sanaa was a confident student with a more serious, professional demeanor than many of her peers. Like Isa, she read on grade level. She had also exited ELL

services, having been at ELPA Level 4 in kindergarten, Level 3 in first and second grade, and Level 1 in third grade. Sanaa regularly participated in class during the argument unit, raising her hand to answer Mr. J.'s questions about science content, language choices, and stages. When she volunteered to read, she did so fluidly and expressively.

Sanaa's interview was generally strong. When asked to identify the stages in her own writing she responded quickly and confidently, though in some instances she misidentified stages in her writing. She was able to talk easily and at length about her use of modality in several cases, even explaining how the meaning may have been different had she made another choice. When asked if she thought her argument would be effective, she noted that some things she'd included would be persuasive, and others (those about which she was not sure herself) less so. One of the most interesting aspects of her interview was that her understanding of the genre allowed her to see where she had left out certain elements or hadn't explained herself well. Her disappointment when she realized her mistakes was evident, but those very realizations highlighted an important point: it is possible for a student to have a fair amount of explicit genre knowledge, as Sanaa's classroom participation and interview responses suggested she did, and for that knowledge to not translate into highly effective writing.

Sanaa's original argument is shown in Figure 5.2 below. The transcribed version reads as follows:

Cane toads are a huge problem in Australia. They brought 102 toads to solve a problem but they became an even bigger problem themselves. There are now 2,000,000,000 cane toads in Australia. Scientists think that the meat ants are a good solution to stop the cane toad invasion. I think that scientists should use the meat ants

because they are probably the only insects who can kill the cane toads. You might think that meat ants are dangerous, but the toxins from the cane toad cannot affect the meat ants. So if I lived in Australia I would not have to worry about getting poisoned. From my point of view scientists should use the meat ants, now!

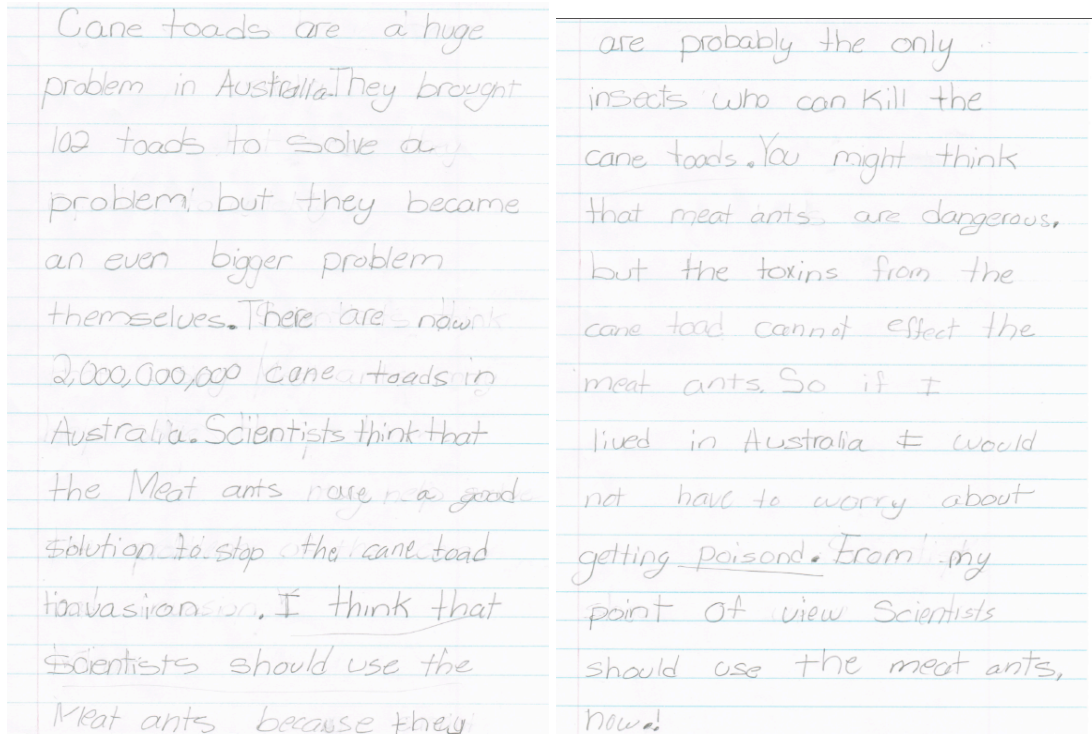


Figure 5.2. Sanaa's handwritten argument

Sanaa's Description of Problem/Claim pair was marked as Successful. She gives background information establishing that cane toads are a problem, that overpopulation is part of that problem, that scientists think meat ants are a good solution, and that she agrees with them. Sanaa's writing at this stage pair was more typical than Isa's in that most students formulated their Description of Problem stage with minimal use of personal pronouns, and many broke with this pattern to introduce their Claim in the first person. The similarities between her Claim and those of many of her classmates

contributed to the analytical decision to label “I think that scientists should use the meat ants” as Claim and to separate it from the rest of the sentence. Although Sanaa read this whole sentence as part of her Description of Problem/Claim, functionally the last part is characteristic of Evidence.

The clause “because they are probably the only insects who can kill the cane toads” composed the entire Evidence/Reason stage pair. There was no Reason whatsoever, which Sanaa realized as she discussed her writing in the interview. The absence of this stage led to the classification of the stage pair as Unsuccessful. The Evidence did accurately reflect the source material and Sanaa was intentional in her choice of “probably” to communicate that she felt she needed more information to know whether meat ants were truly the only insects that could kill cane toads. The Evidence did support her Claim, but it was impossible to establish consistency between Evidence and a nonexistent Reason. There was not enough substance to evaluate whether this part of the argument was logical, or for the stage pair to achieve its intended goals.

Sanaa’s Counterargument and Evaluation were also Unsuccessful. The statements carrying out the respective functions of those two stages were again combined into one sentence. The first part of that sentence, which was effectively the Counterargument, signaled the presentation of someone else’s point of view (“You might think that meat ants are dangerous”). The following clause began with the connector of contrast “but”, signaling Sanaa’s rejection of that point of view—in other words, her Evaluation, which employed the “direct rejection via facts” strategy (“but the toxins from the cane toad cannot affect the meat ants”). This was followed by the comment that if Sanaa lived in Australia she “would not have to worry about getting poisoned,” a statement that does not

fall neatly into any of the functional categories discussed here, though the use of a conditional hints at it being intended as an Evaluation (and Sanaa identified it as such).

Unclear logic and a dearth of factual detail were serious issues for this stage pair. The Counterargument presented is more emotional than scientific, and would clearly not be the strongest point in support of an anti-meat ants view; nor does the Evaluation adequately address the concern raised in the Counterargument. Accordingly, the Logic, Factual Correctness, and Consistency dimensions of the evaluative framework were marked as Ns. Logically, the question of what is dangerous for whom needs to be clarified. The Evaluation and second sentence seem to address the notion that the cane toads themselves are dangerous. At my prompting, Sanaa elaborated on these points in her interview, but though she spoke clearly, her responses did not make any more sense of these statements and I came away with the impression that she had an incomplete or incorrect understanding of the facts surrounding the issue of toxins. The Function dimension received an S: the overall shape and function of the stage pair were recognizable, but the odd second sentence and the serious problems described above affected their ability to “do” what they were supposed to. Language Features was the only dimension to receive a Y in this case; in a classroom context this would be an achievement to point to, but ultimately the value of appropriately using language features is diminished if the overall message has been compromised to such a degree.

Finally, Sanaa’s Restatement of Claim was evaluated as Partially Successful, as was the writing of most of her peers at this stage. The stage was comprised of a single sentence that simply reiterated Sanaa’s position, without any summarizing or additional detail. She maintained the first-person perspective that was present throughout her

argument in the interpersonal theme used as her transition to this final stage (“From my point of view”), and moved to motivate her reader by adding a sense of urgency to her recommendation (“now!”). Overall, this was a basic Restatement of Claim that was perfectly acceptable, though not exceptional.

Where Isa’s strong understanding of the content in the argument unit translated into an effectively written argument, Sanaa was an otherwise strong student from whom we might have expected a strong performance but for whom the understanding reflected in her in-class responses did not translate to her writing. The final focal student, Amir, also struggled somewhat, but bridged the gap between classroom examples and his own writing in some interesting and unexpected ways.

Amir. Amir was a small child with a confident and affable personality. He was in the Approaching Grade Level reading group and scored at an ELPA Level 3. Amir’s ELPA results had remained relatively steady throughout the years: he had scored at Level 2 in kindergarten, gone down to Level 3 in first and second grade, and moved back to Level 2 in third grade, which meant that his fourth grade score was lower than it had been a year earlier. In class he was attentive but quiet, reading with some hesitation when called on but not volunteering often. Limited participation notwithstanding, his teacher counted him among the students who were interested in and motivated by Functional Grammar, and he did remain engaged with the lesson even when just observing. In his interview, Amir was able to quickly and accurately answer questions regarding his choice of “likelihood words”. He seemed to have a fairly good sense of the argument material overall, but it is difficult to say whether his facility with identifying some stages

were attributable to his genre knowledge or to his having labeled them in his text. Amir's original text is shown in Figure 5.3. The transcribed version reads as follows:

The cane toad invasion has become a big problem. The problem is that cane toads have poison in their skin and [they are]²⁸ spreading over Australia. Meat ants are a good solution but the scientists have to study it more.

The ant is a dominant ant. And it can kill the toads. <-Evidence

They are a good solution because they help reduce the toad population. <-Reason

The meat ants fight other ants for their territory. Because if other ants eat the meat ants there will be less ants and there will not be enough ants to kill the cane toad. The part I agree in the counterargument is that ants fight other ants for their territory. How much do I agree? 5. The part I disagree is that they fight other ants for their territory. How much do I disagree? 7. I disagree because other ants fight the meat ants for their territory, and if they eat the meat ants there will be less population of the meat ants and if there is less population of the meat ants there will be more cane toads because the meat ants eat the cane toads.

I think scientists should study meat ants more because the meat ants are killing the cane toads but the scientists can study the meat ants more. And scientists can try it before it goes to living.

²⁸ "They are" not included in original text.

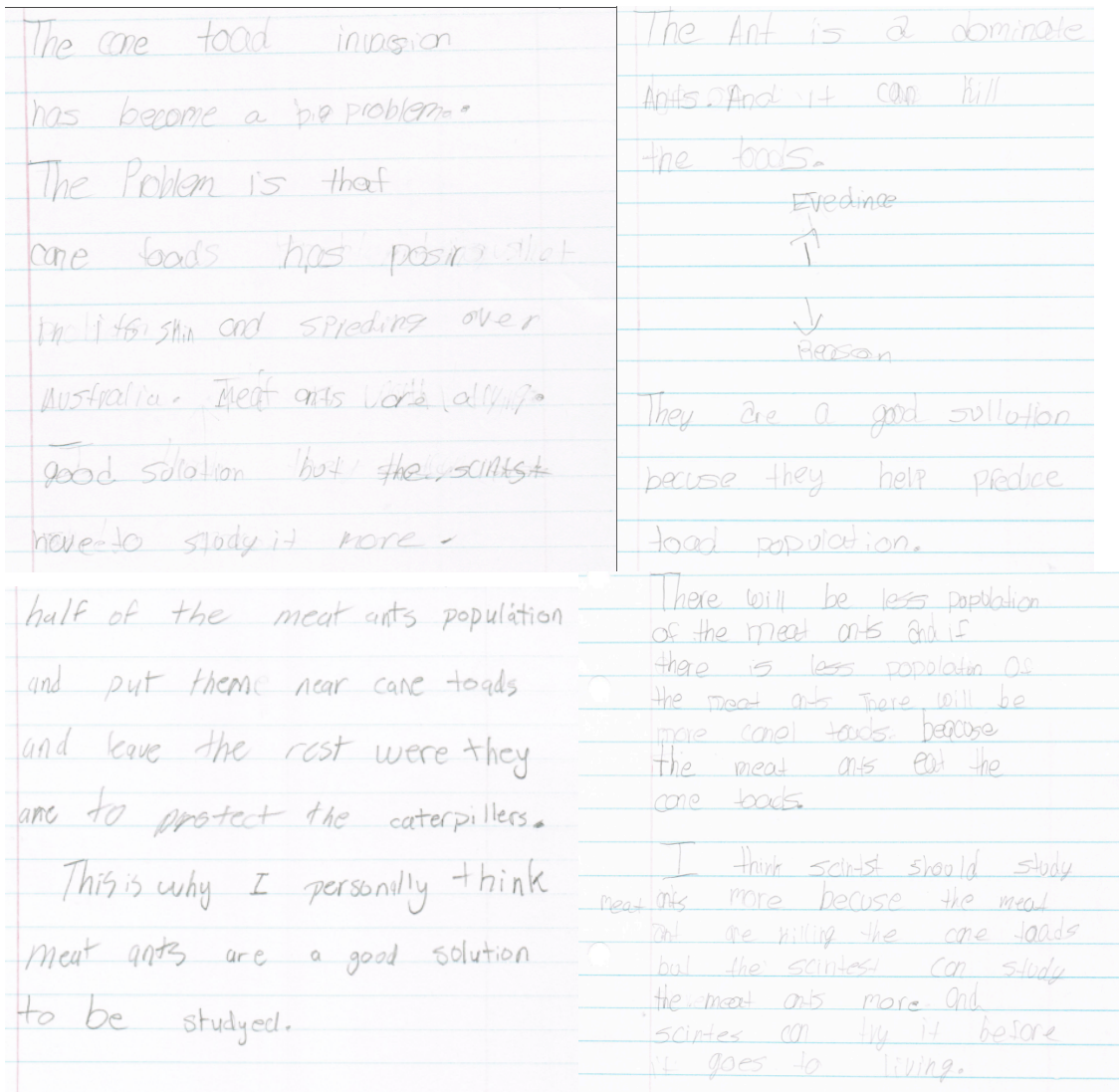


Figure 5.3. Amir's handwritten argument

As is immediately evident, Amir's written response is marked by heavy use of scaffolding. His Evidence and Reason are both labeled, with arrows pointing to the relevant sentences, reminiscent of the deconstruction activities done during the argument lessons. His Counterargument and Evaluation are shaped by that graphic organizer to a much higher degree than any of his classmates. Even the initial and final stages, for which no graphic organizer was provided, seem to have been constructed in lockstep with the descriptions provided for each stage.

In some ways this proved to be a useful strategy. Amir's Description of Problem/Claim pair met the minimum standard for a Successful response: it established that cane toads were a problem for Australia, explained why, introduced meat ants as a possible solution, and presented the writer's position on the issue. It is not as elaborated as some of the strongest examples, but as a basic response it hits the necessary marks.

Amir's Evidence and Reason were Partially Successful. He included two pieces of Evidence, the second of which was adequately linked to the Reason that followed it. Again, this is a very simple argument, but on a very basic level it is logical and factually correct. Both of these dimensions were marked as Ys. The Features, Consistency, and Function dimensions received Ss. There was some transitional language at the beginning of the Reason and Amir indicated that he intentionally engaged with the modality of ability ("can") elsewhere in his argument, so it is reasonable to imagine that the same was true at this stage. In terms of Consistency, Amir's argument is another representative example of most students who chose the "good, study more" claim; his Evidence/Reason pair provides support for the claim that meat ants are a good solution but does not say why they should be studied more. Finally, for the Function dimension, the second piece of Evidence and the Reason are, minus the labeling, appropriately structured. The first piece of Evidence, though, was extraneous if interpreted in terms of its actual meaning. There is another possible interpretation which, if true, would be the first hint of Amir's apparent difficulties with the science content. He may well have understood the word "dominant" in a more everyday sense, taking it to suggest that meat ants would be victorious in altercations with any organism (including cane toads), rather than the more specific scientific sense in which it was used in the source texts, referring to the meat

ants' relationship with competitor ant species. For evaluative purposes, the misunderstanding cannot be assumed, but in a classroom application marking the dimension as an S in light of the seemingly extraneous evidence effectively flags the issue for a discussion in which any questions can be clarified.

The Counterargument and Evaluation suggest more strongly that Amir did not understand the science content well. At the very least, he narrowed his focus in this stage pair to the interaction between meat ants and other ant species to a degree that detracts strongly from the argument's effectiveness. The pair was marked as Unsuccessful, and received Ns for Logic, Factual Correctness, and Consistency and Ss for Function and Features. The latter markings result from Amir's close adherence to the graphic organizer; a response that retains as much of that structure as this one does cannot help but to bring some of the appropriate shape and language along with it. The content of this stage pair, however, is extremely confusing and difficult to parse. There is an identifiable train of thought to do with not wanting other ant species to compromise one of the only viable solutions to the cane toad problem by killing the meat ants, but the actual trail of who argues what and who disagrees with whom is impossibly tangled, and in any case is not derived from the source texts.²⁹ When Amir read me his Counterargument, I asked him who would have said that, and he responded: "Me. Cause I'm telling them bad don't use because bad don't use, they're telling me that some ants fight other ants for their territory, but I'm disagreeing with them, I'm doing a counterargument with them." Amir understood that at this stage he needed to engage someone with a different point a view, but both in writing and in conversation the specifics of how to carry this out seemed to

²⁹ The only information provided in the Meat Ants text on this point reads: "Meat ants are a dominant type of ant. Some ants try to avoid the meat ant. However, some species of ants fight with meat ants to establish territory."

elude him. The fact that he struggled so much with this stage pair is representative in its own way: in addition to the sheer proportion of students who wrote Unsuccessful Counterargument/Evaluation pairs, in interviews with focal students I heard repeatedly that this was the most challenging aspect of the unit.

Amir's Restatement of Claim, meanwhile, was evaluated as Partially Successful, but for a slightly different reason than most other examples that received this designation. Unlike many of his classmates, it is clear that Amir made an attempt to summarize parts of his argument in addition to restating his claim. What he wrote, however, is more repetition and circular reasoning than summary. The infelicity in the last sentence ("before it goes to living") does not hinder effectiveness so much as the intended meaning does—it is ultimately just another way to say that scientists need to study more ("try it") before making changes in the wild.

Amir was a student who, based on his reading group and ELPA placements, one might expect to have a harder time writing an effective argument than students on the other end of the academic spectrum. And it is clear that some aspects of writing an argument were challenging for him. What is interesting about Amir's case is his strategy for overcoming these challenges. In an instructional context designed to be highly supportive of genre learning, Amir is a student who took full advantage of the tools offered to him, hewing closely to what graphic organizers and other classroom artifacts said about what was expected of him. Where Isa internalized and improvised on this support and Sanaa perhaps did not follow it closely enough, Amir embraced it, and the way he did so provides valuable information about the strengths and limitations of our instructional design.

Traditional Writing Assessment

Students' arguments were also evaluated using a primary trait rubric by independent raters—fourth-grade teachers who worked in the Oak Grove district but who had not participated in the FG program. (Chapters 3 and 4 provide additional details about rater selection and scoring procedure.) Exact and adjacent agreement for these arguments in the initial round of scoring was 85% on Ideas & Content, 78% on Organization, 71% on Voice, 92% on Word Choice, 78% on Sentence Fluency, and 97% on Conventions (see Appendix N for more detail).

The scores provided by the independent raters suggest that the stage-based evaluation of arguments used for this study is compatible with more traditional forms of assessment and support a number of findings from the overall analysis. One example of this is difference in how arguments were scored based on the position the author took (see Table 5.8, below). Students who chose the harder-to-support “good, study more” position were rated at least half a point, and sometimes as much as a point and a half, lower than students who argued directly for or against the use of meat ants. It is possible that average scores for the latter positions (especially “bad, don’t use”) are inflated somewhat because of small sample sizes, but the available data show that students’ difficulties in adequately supporting both parts of that claim were reflected not only in the stage analysis but also in how the independent raters viewed their writing.

	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
Position						
"Good, study more" (n=18)	3.30	3.08	3.05	3.33	3.10	3.18
"Good, use now" (n=6)	4.06	3.56	3.42	3.58	3.58	3.42
"Bad, don't use" (n=2)	4.92	4.34	4.67	4.50	4.75	4.75

Table 5.8. Average scores given by independent raters by claim

Comparing the average scores for students at different stage-based performance levels provides additional evidence that such an approach could be integrated with existing forms of assessment. Table 5.9 shows the average scores assigned by independent raters organized by stage or stage pair. Students whose writing was rated as Successful at the Description of Problem/Claim, Counterargument/Evaluation, and Restatement of Claim stages received the highest scores almost uniformly.³⁰ Likewise, students with Partially Successful designations received better scores from independent raters than did those whose writing had been classified as Unsuccessful. On average, the difference between Successful and Partially Successful writing on these stages was about a third of a point on *Ideas and Content*, a quarter to half a point on *Organization*, and a third to half a point on *Voice*. When comparing Successful and Unsuccessful stages these numbers increased to roughly half a point to a whole point average difference. Because different stages within the same argument were almost always rated differently in the stage analysis, this cannot be a matter of the same individuals determining subgroup scores across stages. It appears, then, that effectiveness in crafting the stages does correspond with the kind of writing that is valued in the traditional assessment.

Interestingly, the relationship between stage-based analysis and primary trait scores was the strongest for the Description of Problem and Claim stage pair. The difference between Successful and Unsuccessful writing on the *Ideas and Content* dimension was 0.77 points, as compared to 0.39 at the Counterargument and Evaluation stages and 0.54 for the Restatement of Claim. On the *Organization* dimension these differences were 0.69, 0.19, and 0.53 points, respectively. This may mean that effectively

³⁰ With the exception of Counterargument/Evaluation, where there were slightly higher scores in Sentence Fluency for Partially Successful and Unsuccessful students (0.17 and 0.5 points, respectively).

setting the reader up to follow the argument is even more important than we had imagined in designing the argument unit. It could also be that the presence of a solid “introduction” and, to a lesser extent, “conclusion” signals to teachers that they are reading the work of a competent writer—and by the same token, it could well be that generally strong writers are skilled at producing these more universal stages, even if they were less successful in writing other stages more specific to this argument genre.

	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
Description of Problem/Claim						
Successful (n=12)	3.94	3.61	3.65	3.80	3.79	3.82
Partially Successful (n=10)	3.65	3.37	3.28	3.44	3.24	3.26
Unsuccessful (n=4)	3.17	2.92	2.75	3.25	2.92	3.00
Evidence/Reason						
Successful (n=5)	3.67	3.17	3.50	3.67	3.73	3.80
Partially Successful (n=8)	3.69	3.44	3.50	3.81	3.40	3.44
Unsuccessful (n=10)	3.57	3.25	3.11	3.36	3.24	3.25
None (n=3)	4.17	4.01	3.83	3.67	3.83	3.83
Counterargument/Evaluation						
Successful (n=4)	4.13	3.75	4.00	3.88	3.50	3.63
Partially Successful (n=5)	3.83	3.23	3.50	3.73	3.67	3.67
Unsuccessful (n=11)	3.74	3.56	3.44	3.62	3.55	3.56
None (n=6)	3.00	2.78	2.50	2.97	2.86	2.83
Restatement of Claim						
Successful (n=5)	3.87	3.53	3.60	3.93	3.87	3.97
Partially Successful (n=19)	3.61	3.32	3.28	3.43	3.29	3.29
Unsuccessful (n=1)	3.33	3.00	2.67	3.67	3.33	3.33
None (n=1)	3.00	2.67	2.67	3.33	2.67	3.00

Table 5.9. Average scores given by independent raters by stage

Oddly, the weakest relationship observed between primary trait scores and stage analysis was for the Evidence and Reason pair. Partially Successful writing was rated essentially the same as Successful writing, and there was just a 0.10 point difference between Successful and Unsuccessful writing. This was unexpected, as the Evidence and Reason stages are essential in building a strong argument. The most likely explanation for this observation is that it reflects the difference between the very specific expectations for writing established by this FG approach to argumentation and the demands of the persuasive writing that elementary school teachers are most accustomed to teaching and evaluating. In my stage analysis, a missing Reason would hinder the stage pair's effectiveness enough to merit the Unsuccessful classification. If there is no corresponding expectation of explicit linking of evidence and claim in typical persuasive writing, that absence would easily go unnoticed. The simple presence of reasonable evidence would presumably be sufficient for the student to be perceived as meeting the goals of the genre, and the writing would be evaluated more positively than was the case in stage analysis.

Language and academic proficiency levels had a stronger link to scores given by independent raters in fourth grade than they did for second graders. Table 5.10 presents average scores for these subgroups. The group of students who had exited the ELL program performed better on all traits than those who were still classified as ELLs. Likewise, the students at ELPA Levels 1 and 2 received higher average scores than those at Level 3 on almost all traits. Reading levels broke down into two bands, with students who were reading at or above grade level scoring an average of a point or more higher across traits than students who were below or approaching grade level. Still, these average scores represented a range of individual performance within each group, as with

the stage analysis, in which the higher language and academic proficiency groups had higher proportions of Successful examples but nonetheless included a number of students performing at the Partially Successful and Unsuccessful levels on any given stage. In other words, while there appears to be some relationship between language proficiency/reading level and argument quality, the former does not strongly determine the latter. The difference in scores between the lowest two language and reading groups was generally about two tenths of a point, smaller (in some cases dramatically so) than the difference between these and the higher performing groups; this again raises the possibility that the performance of students who typically struggled was bolstered by the instruction and stronger than it otherwise would have been, though it is difficult to isolate the influence of pedagogy from other factors (for additional discussion of this point, see Chapter 6).

	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
ELPA Level						
Monitored (n=12)	4.22	3.90	4.01	4.03	4.11	4.15
1s & 2s (n=5)	3.27	2.87	2.87	3.00	2.67	2.63
3s (n=9)	3.09	2.87	2.65	3.17	2.87	2.89
Reading Level						
Above Grade Level (n=5)	4.17	3.77	4.00	4.10	4.10	4.30
On Level (n=9)	4.20	3.89	3.91	3.91	3.85	3.83
Approaching Level (n=4)	3.09	2.67	2.75	3.17	2.67	2.75
Below Grade Level (n=8)	2.98	2.82	2.52	2.94	2.83	2.75

Table 5.10. Average scores given by independent raters by subgroup

Some of the parallels between primary trait scoring and stage analysis are highlighted by the scores assigned to focal students' arguments. Independent raters gave Isa's argument 4s in Ideas & Content, Organization, Word Choice, Sentence Fluency and

Conventions, and a 4.5 in Voice. One rater commented “interesting read; voice evident.” Stage analysis indicated that the overall effectiveness of Isa’s argument was higher than average, and the voice features were unique among writing samples in the class, so these results are not surprising. Amir’s argument, meanwhile, received 3s in Ideas & Content and Word Choice; 3.33s in Organization and Voice; and a 2.67 in Conventions. A rater commented that his writing was “not focused; sentences are randomly strung together.” And indeed, lack of coherence and clarity were also issues in Amir’s argument from the perspective of stage analysis. Interestingly, Sanaa’s argument, which had some serious logical flaws, was given a 5 in Ideas & Content and 4s on the remaining five traits. In comparing the two arguments, Sanaa’s handwriting and spelling are both noticeably better than Isa’s, and her writing is more fluid. That her scores were not negatively affected by the substantive weaknesses discussed in stage analysis points again to the different values undergirding the two analytic approaches: while the 6 + 1 Traits rubric includes a nominal focus on ideas and content, it would appear that surface-level features play a stronger role in shaping raters’ perception of writing quality.

The next chapter synthesizes this study’s findings on the characteristics of second- and fourth-grade writing and provides a discussion of the strengths and limitations of the instruction and the study overall.

CHAPTER 6

Discussion

This study has presented careful analysis of argumentative writing collected in the second and fourth grade and the context in which it was produced by providing detailed descriptions of how instruction unfolded over the course of the argument unit in each grade; presenting fine-grained linguistic analysis of student writing at the level of whole class, academic and language proficiency subgroups, and focal students; and discussing results of independent scoring of students' arguments using a traditional primary trait rubric.

This final chapter presents a synthesis of second and fourth grade findings that enable us to see differences in language choices across grade levels as well as providing a comparison of how effectively students in each grade were able to construct the particular kinds of meaning associated with each stage of the argument genre they learned about. By making visible the range of performance at both grade levels, this comparison also allows us to draw implications for the specific areas in which students in each of the two grades may need additional support.

In discussing these findings, this chapter argues for reconsideration of our assumptions regarding the argumentative abilities of young students in light of the features of the writing these students produced with the aid of carefully-scaffolded instruction, and offers concrete pedagogical recommendations for strengthening future

instruction in written argumentation at the elementary level. The chapter concludes with a discussion of the limitations of the study and directions for further research.

Clause-level Analysis

Analysis of second- and fourth-grade texts revealed a number of differences, many of which were expected based on the students' respective developmental stages. For instance, clause-level analysis showed a number of basic differences in overall length and complexity between the grade level groups (see Table 6.1). Fourth graders used almost three times as many clauses on average as did second graders (28.35 versus 10.52) and just over twice as many T-units (14.35 versus 6.12). The average number of embedded clauses for the fourth-grade group was over eight times as large as the average number of embeddings for second graders (6.50 versus 0.80). Measures for clauses per T-unit and embedded clauses per T-unit showed similar growth, consistent with expectations that children's writing increase in complexity as they grow older and advance through school (e.g., Christie, 2012; Hunt, 1977; Malvern, Richards, Chipere, & Durán, 2004). There were also much wider ranges for individual performance on all measures in the fourth grade. These ranges reflect a general increase in diversity in the fourth-grade texts. While we might expect that the range of performance across students would increase as they move through the grades, this increase can also be attributed at least in part to the less constrained nature of the fourth-grade writing task. Students at this level were capable of producing more sophisticated writing and were expected to have less trouble managing numerous stages, and the level of scaffolding built into the pedagogy was adjusted accordingly.

Analysis of student writing in light of the three SFL metafunctions—the ideational, the interpersonal, and the textual—showed that students who participated in this study made linguistic choices that were largely consistent with both the demands of the argument genre and the kind of writing that is typical for writers of their respective ages. Ideationally, writing in both grades was characterized by the use of generalized participants. Students used mostly being and doing processes to present factual information about grizzly bears, cane toads, and/or meat ants. The use of saying processes was appropriately limited, usually appearing (if at all) only in the Counterargument stage, where they were used to report on another party’s point of view. Sensing processes were also used infrequently. Second graders used sensing processes to project their thoughts slightly more often than they used saying processes, while fourth graders used sensing process less frequently than any other process type. This suggests the possibility of a developing awareness that expectations for how “present” a writer is vary across genres. Older students’ writing displayed more sensitivity to the lack of sensing processes and personal pronouns in the model arguments discussed during the unit, though it is not clear whether this was for developmental reasons or by virtue of having been exposed to a wider variety of text types over their two additional years of schooling.

There was an even more marked change between grades with respect to the types of conjunctive meanings students used. Not only did the categories of conjunctive meanings observed in the data double from second to fourth grade, but the way these resources were used also changed. The use of connectors of contrast and addition multiplied severalfold. Connectors of addition jumped from just 5.97% of total

conjunctive meanings in second grade to 22.78% in fourth, contributing to increases in sentence complexity and text length: fourth graders needed to package and link a larger number of ideas, and increased use of “and” or “also” helped them do so. Connectors of contrast, usually “but”, showed similar growth in usage, from 2.99% in second grade to 21.62% in fourth. While second graders needed to draw on contrastive meaning at just one point in their arguments—their Evaluations—fourth graders were more likely to use “but” to engage their readers or the opposition by acknowledging and addressing other points of view throughout their arguments. This increase was also partially driven by the fourth-grade curriculum itself, where the concept of unintended consequences motivated the use of connectors of contrast (e.g., “Scientists brought cane toads to Australia to solve a problem, but...”).

Clauses	Mean	Min	Max	Range	SD
2 nd	10.52	7	13	6	1.39
4 th	28.35	14	45	31	9.56
# T-Units	Mean	Min	Max	Range	
2 nd	6.12	4	9	5	0.97
4 th	14.35	6	21	15	4.85
# E.C.s	Mean	Min	Max	Range	
2 nd	0.80	0	3	3	1.04
4 th	6.50	2	13	11	3.17
Clauses/T-Unit	Mean	Min	Max	Range	
2 nd	1.73	1.33	2.20	0.87	0.29
4 th	2.00	1.53	2.80	1.27	0.20
E.C.s/T-Unit	Mean	Min	Max	Range	
2 nd	0.14	0.00	0.50	0.50	0.17
4 th	0.47	0.15	1.17	1.02	0.23

Table 6.1. Clause-level descriptive statistics for second- and fourth-grade writing

Other types of conjunctive meanings decreased between grades. Connectors of cause or reason accounted for 64.18% of conjunctive meanings in second grade, but their proportion dropped to just 16.99% for the fourth-grade class. For second graders this type of meaning was typically realized in the form of “because.” Fourth graders also used “because,” but, having developed other strategies for conveying causal meaning, no

longer depended on it as heavily. As grammatical metaphor develops, writers begin to replace sentence constituents—in this case, causal connectors—with nominalizations (Halliday, 1994). Fourth graders in this sample, for example, frequently used constructions like “One reason for this is...” in lieu of “because.” In some cases fourth-grade writers also omitted connectors by introducing implicit causality through the juxtaposition of sentences, though this was less common.

A less drastic drop in usage was observed for connectors of condition, from 26.87% of total conjunctive meanings in second grade to 11.58% in fourth. For both grades instances of conditional connectors tended to be the “if” associated with making predictions; the proportional decrease in this case does not mean that fourth graders have developed beyond this type of meaning, but is instead likely attributable to increased usage of other types of connectors. Fourth graders used all of the conjunctive meaning types found in the second-grade data and added meanings associated with time/sequence, purpose, concession, and consequence. The expansion of these resources from second to fourth grade is typical of children’s writing development (Christie & Derewianka, 2008), and within the context of this study, such development allowed fourth graders to manage a task that was more demanding in terms of both scientific content and rhetorical demands, permitting them to build factual explanations, expand on their predictions, and negotiate the various perspectives relevant to the issue about which they were writing.

Interpersonal resources were employed in accordance with the expectations of the genre at both grade levels. In terms of grammatical mood, the overwhelming majority of clauses were declarative; a handful of students at each grade level used interrogative and/or imperative clauses to present rhetorical questions or otherwise add emphasis to

their arguments. Modal meanings were used similarly in second and fourth grade. Modality of obligation, foundational to making a Claim (e.g., “should”), was used by almost all students. A majority of students in both grades also used modality of likelihood, frequently associated with making predictions (e.g., “will,” “might,” “may”). About half of students in both grades used modality of ability (e.g., “The zoo can build a bigger cage”). Though both grades used modality in similar ways, there was a higher incidence of infelicitous usages in second grade, especially with modality of ability; fewer such issues were observed in the fourth grader writing. Finally, although there were no expectations around the textual resource of theme, analysis of this feature confirmed a developmentally expected modest increase in marked themes from second to fourth grade.

With surprisingly few exceptions, neither language proficiency nor reading level were predictive of performance on the measures discussed here. In the second grade, students in different subgroups performed at roughly the same level in terms of text length and complexity. Some differences were noted in relative concentrations of being versus doing processes between groups in second grade, but further analysis revealed that these types of meanings were being used in essentially the same ways; no such pattern was observed in fourth grade. There were slight increases in the total number of conjunctive meanings in second-grade writing as language and academic proficiency levels increased, and a similarly small increase in use of connectors of cause/reason for students who read at higher levels, suggesting a trend toward increased clause complexity and sensitivity to the particular types of meaning valued in a given genre. In the fourth-grade class, use of contrastive resources, which played an important role in the

presentation of information central to many students' arguments, also rose slightly as language proficiency level increased. These numbers are small enough, however, that they must be interpreted very cautiously.

The only other potential relationship identified at the fourth-grade level had to do not with numbers, but with the kinds of clauses used by students in different subgroups: students at the highest reading level tended to use embedded clauses to reiterate their claims throughout their arguments, while those in the lowest two reading and language proficiency groups were significantly more likely to use projecting clauses (e.g., "I think...") that gave their arguments a more personal tone. As neither of these points was directly addressed in instruction, these results suggest differing levels of sensitivity to the purposes and/or language features of the genre, an observation that provides additional support for the necessity of explicit instruction. Higher-performing students appear to be better positioned to absorb more information about the genre than what is explicitly focused on by the teacher—the value placed on an objective, impersonal authorial voice, for example, and how to achieve this—while other students rely on their teachers' ability to identify and explain the role of such phenomena. Failure to equip teachers with deep genre knowledge that they can pass on to their students, then, is likely to contribute to the perpetuation of academic achievement gaps.

This last finding also underscores the importance of going beyond frequencies and ratios in analyzing student writing. Numbers can be useful, but can also obscure meaningful information about how students use language to construct meaning. The stage analysis conducted for this study produced several more examples of what such an in-depth examination can reveal.

Stage Analysis

For the purposes of this study, the argument genre was described to students as having seven stages: Description of Problem, Claim, Evidence, Reason, Counterargument, Evaluation of Counterargument, and Restatement of Claim. While the first six stages have distinct purposes and features, they each function in close concert with another stage. These interdependencies were taken into account during stage analysis by taking the stage pair as a unit of analysis in addition to each stage individually. In stage analysis, stages and stage pairs were closely examined to determine how successful students had been in achieving the functional goals of each stage. This analysis allowed for detailed characterization of students' first attempts at writing this genre as well as identification of areas in which pedagogy could be strengthened to provide additional support targeted to the points with which some students were likely to struggle. The following section reports the overall findings of this analysis. A summary of performance at each grade level is found in Table 6.2.

	Second Grade	Fourth Grade
Description of Problem and Claim	Co-constructed, not analyzed in detail	Students successfully stated Claims; deciding how much or little detail to include in Description of Problem presented a challenge
Evidence and Reason	Selection of Evidence was unproblematic, but some students struggled to make logical connections clear in their Reasons	Determining appropriate scope for Evidence a problem for some; overall students were better able to explain Reasons, but many had difficulty adequately supporting a "middle ground" position
Counterargument and Evaluation	Careful scaffolding of Counterargument enabled success at this stage, but performance on Evaluation was strongly determined by the fact chosen as the Counterargument, as students did not have language to effectively frame responses to Counterarguments that could not be directly rebutted	Students were largely successful in presenting basic Counterarguments, but most would benefit from further elaboration; a broader range of strategies for Evaluation was used, the effectiveness of which determined success for this stage pair; linguistic resources for concession and contrast still limited
Restatement of Claim	Word-for-word restatement of initial claim	Basic rewording of claim was standard; few students added summary of position

Table 6.2. Grade level differences in performance

Description of Problem, Claim, and Restatement of Claim. Classroom materials given to students defined Description of Problem as a stage that *“familiarizes the reader with important background on the issue or problem.”* Claim was described as *“a statement that introduces the specific issue to be argued and the writer’s position on that issue.”* The description of the Restatement of Claim stage provided on instructional materials was *“Summarizes and restates the writer’s position on the issue.”* At the second-grade level, Description of Problem, Claim, and Restatement of Claim were not analyzed in-depth. Because adequately summarizing the background information required in the Description of Problem was expected to be a challenge for second graders, this stage was co-constructed during one of the argument lessons and was shared by all students. Claim and Restatement of Claim generally consisted of the same short phrase at the beginning and end of second graders’ arguments—“We should (or should not) put grizzly bears in zoos”—and only minor variations were noted between arguments.

At the fourth-grade level, Description of Problem/Claim was the stage pair for which performance was best overall. A total of 26 students included Description of Problem and Claim in their arguments, of which 12 students’ pairs were categorized as Successful, 10 were Partially Successful, and just four students were Unsuccessful. As in the second grade, Claim was a straightforward stage, consisting of a single sentence; the only instances in which a problem was noted were the few in which a writer did not include an initial Claim but either inserted it elsewhere in the argument or made a clear Restatement of Claim in conclusion. The Restatement of Claim stage was similarly unproblematic, if also a bit uninspired: while almost all students restated their position, only five did anything more, suggesting a need to emphasize the expectation that this

stage also include a brief summary. Still, these students appear to be practiced enough in stating their position on an issue that doing so in this argument genre posed little challenge overall.

The Description of Problem stage, on the other hand, revealed that what was expected to be difficult for second graders was also challenging for many fourth graders. Problems with this stage turned primarily on issues of scope: how much or how little detail to include. At minimum, a Successful fourth-grade response would have to establish that cane toads were a problem for Australia, explain why, and introduce meat ants as a possible solution. Even the students whose writing satisfied these criteria tended to have minor issues with finding the right balance of information. For those who did not have a well-developed ability to anticipate what a distant reader would need to know, such issues were magnified.

In terms of adjusting future pedagogy to better support this latter group, only one specific language issue was identified; namely, the difference between *presenting* and *presuming* reference, with the former usually being more appropriate in the introductory stages. This was not an issue with which all students struggled, but a mini-lesson designed to raise awareness of and provide practice on this point would be a fairly simple way of improving the Description of Problem specifically for affected students, and certainly has the potential to benefit writing across genres for all.

In terms of more profound changes to the curriculum, there is a clear need for more attention to summarizing source material and reformulating it for use in the Description of Problem. This could come in the form of guided discussion in which the class worked together to decide which information would be important to include and

what could or should be presented elsewhere. Alternately, students could be presented with the minimum criteria for success at this stage, a pedagogical choice that would represent a middle ground between the strict co-construction approach used in second-grade instruction and the less strongly-scaffolded approach used in this iteration of the fourth-grade curriculum. While the results of this study make clear that some form of guidance is still needed at this level, the question of just how much to offer remains outstanding.

Evidence and Reason. Classroom materials described Evidence as “*a piece of information from the text that supports the overall claim*” and Reason as “*a statement about why/how the evidence supports the claim and interprets the evidence in relation to the larger issue.*” Second graders did a fairly good job with this stage pair: 10 students were classified as Successful, 11 were Partially Successful, and just three were Unsuccessful. Students had encountered and discussed the evidence in the source material a number of times and the class had generated lists of evidence supporting each position. Writing the Evidence stage was largely a matter of selecting a fact and presenting it in the argument. Unsurprisingly, this posed little challenge for the students. The same cannot be said, however, for the Reason stage. The most common issue for this stage pair was difficulty crafting a Reason that made a clear connection between Evidence and Claim. Most of the samples classified as Partially Successful exhibited what I call an incomplete bridge. In these cases, there was enough of a connection between Evidence and Reason to be able to follow the argument, but more of this work fell to the reader than is ideal. Sometimes this was a matter of the logic being mostly clear and simply in need of a clause explicitly acknowledging the connection; in other

cases readers needed to make several logical steps on their own to piece everything together.

Future iterations of the second-grade curriculum might improve upon the current support for this stage by urging students to include the connector of *cause/reason* “because” and incorporating it into model arguments; this would serve as a reminder to strengthen presentation of their logic by thinking about the “how and why” for those who are not otherwise inclined to do so. A number of grammatical infelicities—sometimes serious enough to obscure the author’s intended meaning—were also introduced into the writing through students’ attempts to make predictions in their Reasons using conditional sentences. Although the model texts showed appropriate usage, direct support for this challenging construction was not built in to the argument lessons. A mini-lesson or other activity helping students practice the proper formulation (if + present tense verb in subordinate clause, will or other modal with verb in main clause) is called for if students are to be encouraged to include predictions at this stage.

In making these observations, it is necessary to note that this is one case in which stage analysis may not concord with the way most readers would judge the quality of students’ arguments. Indeed, the facts available to the second graders had been simplified to a degree that even when logical connections were not made explicit it was not difficult to imagine what the writers were thinking. Furthermore, research has indicated that even upper elementary-level students routinely omit information they consider to be shared knowledge (Anderson et al., 1997). Despite indications that some students struggled with writing Evidence and Reason, the results of stage analysis actually give reason for optimism. Analysis of this stage pair revealed that, counter to claims that second graders

are not capable of formulating Reasons, nearly half of the second graders in this study did so successfully. It appears, then, that a reconsideration of younger students' capabilities is warranted: what children may or may not do spontaneously and what they are able to do with careful scaffolding can be two very different things.

The fourth-grade data provide a different kind of evidence that curriculum and instruction play a key role in student performance. Where all but three students in the second-grade class produced Evidence/Reason pairs that were at least partially successful in achieving the goals of those stages, fourth graders did not fare nearly as well; almost half of the class failed to write this stage pair effectively. Of 23 students who included this stage pair, six were Successful, seven were Partially Successful, and 10 were Unsuccessful.

Unlike the second graders, the fourth graders had some trouble writing the Evidence stage. Whereas Evidence in second grade consisted of a straightforward reiteration of a single, simple fact, fourth graders writing this stage were tasked not only with evaluating a larger amount of source material and making judgments about the appropriate amount of factual detail to include but also taking into account not a single characteristic but explanations of systems and/or processes in their reporting of these facts. Though a basic explanation was in reach for most students, the step of analyzing and selecting relevant facts from the source material alone presented some level of challenge for even the highest-performing students. This issue relates to the problem of scope seen in fourth graders' Descriptions of Problem, and it is possible that increased support for that stage would have a positive effect here, too, as it would help students determine the level of detail needed to make a given fact or set of facts most effective in

supporting their positions. It may also be necessary to expand the time allotted to discussion of evidence and make anticipation of readers' needs for information a key point in that discussion. Distilling and appropriately repackaging information is a complex skill set that continues to develop throughout the school years (Graham & Perin, 2007). This of course means that there is no simple pedagogical fix, but there are nonetheless possibilities for improvement.

Another area in which a number of fourth graders struggled was the Reason stage. In general, fourth graders were considerably more skilled in articulating why the evidence they presented supported their position. However, when it came to supporting the claim that meat ants were a good solution that needed to be studied more, many students lost their footing. This position, chosen by some 65% of the fourth-grade class, required writers to demonstrate two things: first, that meat ants were a promising solution, and second, that enough questions remained about their safety or efficacy that they should not yet be used. While it was well within most students' reach to provide support for one of these points or the other, few were able to do both. Tellingly, of the six students who wrote Successful Evidence/Reason pairs, only three had chosen the "good, study more" position. The other three argued either directly for or directly against the use of meat ants. It is possible that a defense of either of these positions could turn out to be less sophisticated scientifically—it seems clear that meat ants are a good potential solution, but Australia's history of ecological disaster brought on by human intervention provides ample reason to be cautious about moving ahead with using them—but there is no question that it is rhetorically easier to support the claims at either extreme. Given the added challenge inherent to taking a more measured position, it may be helpful to delay

introduction of a third, more complicated option as a possible claim. Students should instead be provided with the opportunity to practice supporting simpler claims in their first attempts at writing the genre, advancing to the more challenging positions once a basic foundation is in place. This could take place within the same school year or, in the vein of Bruner's (1960) spiral curriculum, across multiple grades, with students revisiting and adding to their knowledge of argumentation at least once a year.

Counterargument and Evaluation. The Counterargument and Evaluation stages were described in classroom materials as, respectively, "*A piece of information that someone presenting another side of the argument might use for evidence*" and "[*a stage that*] *evaluates the strength of the counterargument and explains what this information means for the writer's position on the issue.*"³¹ These stages were the most challenging for students in this study, as evidenced by both their performance and their comments: when asked what they had found most difficult about the argument unit, a majority of focal students at both grade levels named this stage pair. Still, stage analysis shows that successful execution of these moves is within the grasp of students at these ages, and their struggles reveal aspects of the genre that the research team had not previously considered.

Second graders were more successful in writing Counterargument/Evaluation pairs overall. All 25 students included these stages in their arguments. Eleven wrote Successful pairs, while nine were Partially Successful and five were Unsuccessful. Because of the extensive scaffolding of this task—in the form of multiple discussions, lists on poster paper, and an interactive activity—students were well aware of the points

³¹ The second grade description for this stage pair read: "*Share information that the other side might use as evidence. Why do you disagree?*"

that could potentially be raised by those on the opposing side of the issue. This meant that constructing the Counterargument was simply a matter of choosing the point they wanted to respond to and combining it with the sentence starter “Some people say.” As with Evidence, almost all students did well on the Counterargument itself.

What was most striking about the results of analysis for this stage pair was the degree to which a student’s success at the Evaluation stage—and therefore for the stage pair overall—was determined by the specific Counterargument he or she chose to present, a choice which was in turn shaped by the source texts themselves. Each of the 11 students whose writing was classified as Successful chose the same fact to present as a Counterargument, regardless of their overall position. This was not a coincidence. This fact (that grizzly bears need a sizeable amount of space) was the only available Counterargument that could be directly and logically rejected. Others, like the fact that grizzly bears’ habitat is shrinking due to humans’ encroachment on their territory or that many grizzlies are shot in the wild, were impossible, or at least unreasonable, to totally reject: they were, unquestionably, valid concerns. This difference meant that while the use of “I disagree because...” as modeled in the argument unit was sufficient for students who took on the space issue, other writers needed the linguistic tools to concede a point to some degree but present an overriding concern. The second graders in this sample had not yet refined such skills, and were thus poorly positioned to write effective Evaluations.

The extent to which the facts available would shape performance on the writing task was not something we anticipated in planning or delivering the argument unit lessons. The students’ responses during the interactive counterargument activity hinted at this being more difficult for the second graders than Evidence and Reason, but at the time

we attributed the problems we observed to the level of cognitive challenge the task posed for young students. In constructing the source texts we paid close attention to providing an adequate quantity of evidence for both positions. Only through the analysis of writing data did it become clear that these potential responses were qualitatively different.

This finding has important implications for the way teachers approach the use of informational source texts for argumentative writing tasks. It is unlikely that this challenge can be completely avoided in argument instruction. Attempting to create source texts that bypass the problem would not only likely result in inauthentic material, but would fail to prepare students for the real-world reading they will encounter outside the context of this unit. Instead, it is important to recognize that not all informational texts are created equal: each will have unique affordances and challenges as a source of support for building an argument. Teachers (as well as curriculum and test designers) should include in their planning process a review of the particular texts they have chosen in order to identify instances in which the available facts, in light of the specific prompt to which students will respond, may pose the type of challenge discussed here. Students, in turn, need to be supported to recognize when a particular rebuttal strategy is called for and provided with language resources to help them carry it out. At the second-grade level, a simple focus on connectors of contrast would likely be sufficient. In addition to the currently suggested “I disagree because”—a form that most strongly supports the direct rejection strategy—teachers could provide practice with language for making a partial concession, such as “I agree, but...” or “That’s true, but I still think [claim] because...”, with the range of connectors (and evaluative moves) to expand as children become more experienced writers.

Performance on the Counterargument/Evaluation at the fourth-grade level was weighted more heavily toward the Unsuccessful end of the spectrum. Twenty students included this stage pair in their arguments, of which just four were rated as Successful and five as Partially Successful. The remaining eleven Counterargument/Evaluation pairs were marked as Unsuccessful.

As in the second grade, most fourth graders were able to identify and present evidence in support of an opposing position, despite a minor oversight in instructional planning. The strongest Counterarguments included both evidence that could be used to support a different position and a Reason explaining why, drawing on the same skills required in Evidence/Reason, applying them to an opposing position, and giving a balanced explanation of why the reasonable counterargument is ultimately not convincing. Though the various facets of this stage were identified and discussed in the course of classroom activities, they were not fully anticipated at the planning stage and were not accounted for in instructional materials used in piloting the argument curriculum. The graphic organizer used to scaffold the writing of this stage pair, for example, did not make note of the need to include a Reason, and neither did discussion prior to the point at which the need for its inclusion was identified. While several students appear to have comprehended this on-the-fly change to the pedagogy, it was simply not as strongly scaffolded as other elements of the genre were. In light of the weaker support provided on this point, the stronger examples of Evidence can be seen as indicative of fourth graders' capabilities, and the average responses must be evaluated not as lacking the Reason element but as complying with the goals of this stage as they were initially communicated.

Just as it had in the second grade, in the fourth grade the Evaluation stage again determined the overall strength of the stage pair. Problems at this level were not driven by the material; not only was there a broader range of material to draw from but the fourth graders were also more skilled in using the basic concessive resources with which their younger peers struggled. Stage analysis identified six distinct strategies used by fourth graders in constructing their Evaluations. These strategies included direct rejection of the counterargument using facts; providing an alternate solution to address the concern raised in the counterargument; direct rejection of the counterargument using a hypothetical; total concession of the point raised in the counterargument; direct rejection based on opinion; and making an unrelated statement. Table 6.3 provides examples of these strategies taken from student writing. Each example in the table begins with the student's Counterargument and is followed by the Evaluation (underlined); the language exemplifying the strategy is italicized.

It was somewhat surprising to see the range of strategies used to formulate Evaluations in the fourth grade, given that model texts and classroom discussion emphasized the factual approach. And indeed, using facts to explain why a given counterargument was unconvincing was the strategy most often employed by Successful writers at this stage. Other types of responses were less likely to be effective, and in some cases may not have been used as intentionally as the word "strategy" suggests. The opinion-based and unrelated statement approaches, which were only associated with Unsuccessful writing, are likely indicative of students struggling to understand how to write in this new genre.

Strategy	Example
Direct rejection of counterargument using facts	Some people may say that we should not use the meat ant because they could fight with other ants. <u>It is true that meat ants can start to fight with other ants for their territory because meat ants are the dominant type of ant. <i>But they do not always fight.</i></u>
Direct rejection via hypothetical	Some people might say meat ants can kill toads. The toxins aren't likely to hurt the ants. The toads just sit there. So the toad population could decrease. <u><i>What if they don't eat the cane toads?</i></u> We don't know if they will eat the toads, so the humongous toad population could get larger.
Direct rejection via opinion	Some people say it's worth trying <u><i>but in my opinion the scientists should study it more.</i></u>
Alternate solution	If someone might argue with me that said it is bad, don't [use] it, they might say the meat ants eat honeydew from caterpillars and butterflies. So it will be hard to make them eat the cane toads. <u>I agree that they eat honeydew, it is their main food. But I disagree that it will be hard to make them eat cane toads. <i>They can put honeydew on the cane toads.</i></u>
Total concession	Some people may say that the ants mostly eat honeydew from caterpillars. They scavenge carcasses. <u><i>This changed my mind because if caterpillars eat honeydew mostly, they may never eat cane toads again. They scavenge carcasses all the time so cane toads are probably the only living things the meat ants eat.</i></u>
Unrelated statement	The meat ants normally eat honeydew from caterpillars and butterflies. The meat ants might eat the cane toads sometimes but normally they eat the honeydew. <u><i>I strongly disagree because most of the meat ants' bites kill the cane toads immediately.</i></u>

Table 6.3. Examples of strategies for Evaluation used by fourth graders

Composing an Evaluation completely disconnected from the Counterargument and attempting to form a rebuttal consisting of nothing more than a reiteration of one's disagreement suggest a lack of understanding of the basic function of one or both stages; the latter also hints at a possible failure to differentiate between a fact-based argument and the opinion-heavy styles of persuasive writing produced in earlier grades. Similarly, it is probable that the instances of total concession (“This changed my mind because...”) were a byproduct of students' interaction with a scaffold that, in order to encourage thoughtful, measured Evaluations, asked them to consider which part of the Counterargument they agreed with and why; in these cases, writers' responses to what was intended to be a thinking tool were prematurely (and incompletely) imported directly into their arguments. It is probably best to treat instances of these three ineffective

strategies as misunderstandings of the genre goals or, in the latter case, the tools provided to help with writing.

Both the alternate solution strategy and the appearance of hypothetical situations or questions, on the other hand, echo remarks made in the classroom at one point or another. Midway through the first lesson, before having read either the lungworms or meat ants texts, students were asked to brainstorm their own solutions to the cane toad problem, which may have planted the seed for this approach. The instructor also used “what if” questions a number of times throughout the unit to help students refine their thinking on a particular question. It is possible that this practice, combined with the emphasis on likelihood and predictions, contributed to the use of this strategy. While fact-based Evaluations of Counterargument were held up as the ideal in the argument unit, the strategies that emerged here also have the potential to be effective. Though incorporating them into instruction may introduce an undesirable level of complication for students just becoming acquainted with this argument genre, these strategies (assuming they are well-implemented) should be considered valid for evaluative purposes.

A pedagogical adjustment that would position students to write the Counterargument/Evaluation stage pair more effectively is the addition of a focus on the language of concession and contrast. Though contrastive and, to a lesser extent, concessive resources were used in the fourth-grade writing, they were quite limited in range. As in the second grade, “but” figured prominently in conveying contrastive meaning, with little variation. Providing models of the way these language resources might be used in writing Evaluations would equip writers with additional options for

expressing themselves, with the added benefit of clarifying the function of the stage for those still in doubt. For example, the most basic responses at this stage simply use connectors of cause: “I (dis)agree because...” A slightly more sophisticated approach would be to use language that partially concedes the counterargument, yet conveys an overriding concern: “It is true that [counterargument]. However...” Similarly, linguistic tools related to contrast and concession could be used to modify responses framed as total concessions to partial concessions: “Even though this is important information...” or “This is a good point. Still...” Such language could even be used to initiate a hypothetical: “I agree that [part of counterargument conceded]. On the other hand, what if...?” An activity modeling these language points and providing practice would be consistent with the FG emphasis on teaching language features in the service of some larger goal. The purpose, then, would not be to include contrastive or concessive language for its own sake, but to aid students in carrying out the function of evaluating the strength of the counterargument and explaining what that information means for the writer’s position on the issue.

At both grade levels, performance on all stages was spread across a spectrum of reading and language proficiency levels. Membership in a particular subgroup was not predictive of how effective a student’s writing would be on a given stage. Nor, for that matter, was their performance on other stages: no student in this study received the same categorization for all stages. Above all, this detailed examination of student writing showed that students’ understanding of and ability to achieve the goals of each stage are highly individualized. The fact that students in the lower academic proficiency groups in particular were able to, counter to usual expectations, effectively compose elements of

the genre does suggest that the pedagogical approach piloted in this study created a particularly supportive instructional environment. Instances in which they and other students struggled despite this support provide valuable input for researchers and curriculum designers hoping to enhance instruction in argumentation for all students. Table 6.4 summarizes the pedagogical recommendations arising from the findings presented in this and previous chapters.

	Second Grade	Fourth Grade
Description of Problem and Claim	Maintain co-construction of Description of Problem as a feature of instruction	<ul style="list-style-type: none"> • Address issue of scope by establishing and discussing minimal standards for success: What would a good response include? What information would work better in another part of the argument? • Raise awareness of <i>presenting</i> vs <i>presuming</i> reference; practice
Evidence and Reason	Support writing of Reason by: <ul style="list-style-type: none"> • Scaffolding use of connector of cause/reason <i>because</i> • Providing practice with formulation of conditional sentences • Maintaining focus on modality of likelihood and ability to support students in making predictions 	<ul style="list-style-type: none"> • Address issue of scope by expanding time allotted to discussion of Evidence and making anticipation of readers' needs for information a key point in that discussion • Limit potential claims to for/against positions for initial introduction to genre, adding more challenging, measured position after basic proficiency is established • Maintain focus on modality of likelihood, ability, and obligation associated with making predictions and accurately reporting on source texts
Counterargument and Evaluation	Expand support for writing of Evaluation by: <ul style="list-style-type: none"> • Drawing attention to different rebuttal strategies and when they might be called for • Provide practice with language of concession ("I agree, but..." or "That's true, but I still think [claim] because...") • Continuing to discuss use of modality in responding to Counterarguments 	<ul style="list-style-type: none"> • Provide continued support for modality associated with making predictions and accurately reporting on source texts • Further develop scaffolding of Counterargument to include both Evidence and Reason in support of opposing position • Add a language focus on expanding resources for concession and contrast to aid in evaluating the strength of the counterargument and explaining what that information means for the writer's position on the issue

Table 6.4. Pedagogical recommendations by grade level

Independent Scoring Using Traditional Primary Trait Assessment

Stage analysis provides a valuable picture of student performance based on a fine-grained assessment of student writing, but it is not widely used by teachers or designers of high-stakes tests. As reported in previous chapters, to determine whether there was any relationship between SFL analysis and more commonly used styles of writing assessment, local second- and fourth-grade teachers who were not current participants in the Functional Grammar program were recruited to rate the arguments produced for this study using a primary trait rubric. The rubric, which had been used in Spring 2012 by all Oak Grove teachers, was an adaptation of the 6+1 Traits Writing Model (Culham, 2005). Under this model, which is used in Oak Grove’s writing program and in school districts across the country, writing is scored with rubrics emphasizing six key constructs: Ideas, Organization, Voice, Word Choice, Sentence Fluency, and Conventions, with a separate focus on Presentation (the eponymous “+1”). Oak Grove adapted the 6+1 Traits rubric for persuasive or “opinion” writing and developed descriptions for each construct, which were then leveled for lower and upper elementary grades.

To determine reliability of the independent scoring, inter-rater agreement was calculated for the initial round of scoring. Combined exact and adjacent inter-rater agreement for the three writing tasks was 88% on Ideas & Content, 85% on Organization, 74% on Voice, 86% on Word Choice, 81% on Sentence Fluency, and 92% on Conventions. (For additional detail, see discussion in Chapters 4 and 5; tables showing exact and adjacent agreement for each task and trait are located in Appendix N.) Independent scores were averaged to produce final scores for each student’s pre- and post-instruction tasks and the argument they had produced over the course of the unit

(“main argument”). These scores were then compared with the results of the stage analysis. The results of this comparison will be discussed below.

Main Arguments. The scores provided by independent raters for the *Grizzly Bears* and *Cane Toads* arguments support the conclusions of the stage analysis and suggest that it is compatible with traditional forms of assessment, although comparison of the respective results makes clear that there are differences in the values that guide each approach. For example, in the second grade, stage analysis demonstrated that it was significantly more challenging to argue that grizzly bears should not be placed in zoos, and indeed, students who adopted this position scored lower on average on all six traits measured by the rubric. Similarly, fourth graders who argued that meat ants were a good solution that needed to be studied more performed more poorly from the perspectives of both the stage analysis and primary trait assessment, receiving average scores for the latter that were half a point to as much as a point and a half lower across dimensions than scores for students who had chosen either of the other two positions. Comparison of the two sets of results along the lines of Claim highlights a key advantage of stage analysis: where the traditional rubric allowed raters to indicate through lower scores that something was amiss, stage analysis was a sharper tool, able to pinpoint the varying degrees of challenge associated with different positions as a likely source of students’ difficulties.

Independent scoring also supported the Successful/Partially Successful/Unsuccessful performance categorizations used in stage analysis. At the second-grade level, for both Evidence/Reason and Counterargument/Evaluation pairs, students whose writing had been evaluated as Successful received the highest average

scores for Ideas & Content, Organization, and Voice; likewise, Partially Successful samples were rated lower than Successful but higher than Unsuccessful writing. The same pattern was observed at the fourth-grade level for Description of Problem/Claim, Counterargument/Evaluation, and Restatement of Claim. While neither of the stage pairs analyzed in the second-grade data appeared to carry more importance for independent raters, the strongest relationship between stage analysis and primary trait scores in the fourth grade was observed for Description of Problem/Claim. This may mean that effectively setting the reader up to follow the argument is even more important than we had imagined in designing the argument unit. It could also be that the presence of a solid “introduction” and, to a lesser extent, “conclusion” signals to teachers that they are reading the work of a competent writer—and by the same token, it could well be that generally strong writers are skilled at producing these more universal stages, even if they were less successful in writing other stages more specific to this argument genre.

Independent scoring of students’ arguments also revealed important differences between stage analysis and traditional approaches to evaluating arguments. Surprisingly, the weakest relationship observed between primary trait scores and stage analysis at the fourth-grade level was for the Evidence and Reason pair. Partially Successful writing was rated essentially the same as Successful writing, and there was just a tenth of a point’s difference between Successful and Unsuccessful writing. This was unexpected, as the Evidence and Reason stages are essential in building a strong argument. The most likely explanation for this observation is that it reflects the difference between the very specific expectations for writing established by this FG approach to argumentation and the demands of the persuasive writing that elementary school teachers are most accustomed

to teaching and evaluating. In my stage analysis, a missing Reason would hinder the stage pair's effectiveness enough to merit the Unsuccessful classification. If there is no corresponding expectation of explicit linking of evidence and claim in typical persuasive writing, that absence would easily go unnoticed. The simple presence of reasonable evidence would presumably be sufficient for the student to be perceived as meeting the goals of the genre, and the writing would be evaluated more positively than was the case in stage analysis.

Written comments collected from the second-grade raters also provided support for the observation that the two approaches to writing assessment differ in terms of the importance they place on surface-level features. While elementary school teachers, who are held responsible for the development of basic handwriting, spelling, and punctuation skills, are understandably concerned with these features, the FG program focuses on ideas and how language is used to convey them, positing that while conventions are not without value, attention can be paid to them once the ideas around which a piece is built are in place. Accordingly, stage analysis did not take these things into account. Arguments were transcribed and spelling and punctuation were standardized for understandability and ease of analysis. Raters' comments, however, indicated that these matters were salient for them. Instructions for raters welcomed comments on whatever they wished to say about a particular writing sample, and there was a notable difference in the kinds of comments they made for arguments that were easy to read versus those that were not. Samples written by students with better penmanship received substantive comments about the quality of their words and ideas; those produced by students with poor handwriting tended to receive comments remarking upon just that.

This phenomenon raises the question of whether the content of students' arguments was treated equally in all cases, or whether factors like readability had an outsized influence on how certain samples were perceived, raising questions about the accuracy of traditional rubrics. If second-grade raters in this study did place more emphasis on surface features, they would not be alone (Chase, 1968; Markham, 1976). Hillocks (2002) observes that standardized tests put a premium on mechanics and organization rather than content. This emphasis is not limited to testing contexts, but reverberates throughout the education system as teachers align their writing instruction to reflect the values of these tests (Coker & Lewis, 2008). It stands to reason that these values would not cease to influence teacher behavior, consciously or otherwise, when it comes to the application of a primary trait or holistic rubric, even one designed to tease apart content and mechanics.

The importance of Reasons in fourth grade and conventions in second were not the only values that differed between stage analysis and the primary trait assessment. Other differences emerged through evaluation of scores assigned for the pre- and post-instruction tasks, to be discussed in the next section.

Pre- and Post-Instruction Tasks. This study's design included pre- and post-instruction argumentative writing tasks intended to gauge students' unscaffolded performance on a prompt similar to the one used in the argument unit before and after participating in the unit's lessons. The prompt was the same for the pre- and post-intervention tasks, and the same prompt was used in second and fourth grades. The subject matter was drawn from the FG unit that students at both grade levels had completed just prior to the argument unit, which featured scientific content focused on

the impact of human activity on different ecosystems. Arguments produced in response to the pre- and post-instruction prompt were evaluated by independent raters following the same procedure used to score the grizzly bear and cane toads arguments (see Chapter 3 for a detailed description of this procedure and Chapter 4 for a description of the prompt and relevant background information; the rubrics used by raters can be found in Appendix B).

A majority of students at both grade levels—80% in second grade and 80.77% in fourth grade—improved on some or all of traits measured by the independent raters' rubric. In both classes, the students whose performance decreased across most or all traits represented a range of language proficiency and academic performance subgroups. Of the 20 second-grade students whose performance improved, 11 were given higher scores on all traits; the same was true for 13 of the 21 students whose performance improved in fourth grade. The remaining students in both grades improved on some traits but their scores decreased for one or more others. About a third of students in this group (seven in each grade) saw a decrease in their Conventions score. In the fourth grade, four students also had their Sentence Fluency score decrease. In both grades there were only one or two instances of lower scores for each of the remaining four traits. That decreases were concentrated in the Conventions category is not a coincidence; as previously discussed, this is consistent with Functional Grammar's emphasis on ideas and how they relate to one another over surface-level features, and is likely reflective of students' focus being allotted accordingly. The appearance of Sentence Fluency as an affected category could be similarly reflective of a focus on including the elements of the genre that had been introduced in the argument unit, to the detriment of the kind of fluidity measured by this

trait—albeit for just four students.

In neither group did there appear to be either “ceiling” or “floor” effects. At both grade levels, the students who performed best on the pre-intervention task (those with who generally had scores of 3.5-4 across traits in second grade, 4-4.5 in fourth) were essentially equally likely to receive higher or lower scores; while some of these initially high performers did see a decline in performance on the post-intervention task, others with similar initial scores were rated even higher on the second task. The lowest-performing students had scores in the 1-1.5 range across traits in the second grade and around 1-2 in fourth. While in a very few cases it would not have been possible for scores to decline, this was not an issue: children who initially scored among the lowest in their classes uniformly showed improvement.

In several cases, it was also the students in the lowest language and reading groups whose writing improved most in terms of the traits that aligned most strongly with the goals of the intervention. Table 6.5 shows the average change in points for each trait in both grades, organized by subgroups. On the Ideas & Content trait, second-grade students at ELPA Level 5 and those in the At Risk reading group made average gains of over half a point, more than any other subgroup. The Level 5 language subgroup also showed the strongest gains for Organization, though the At Risk group made gains close to the class average. In the fourth grade, students at the Approaching and Below Grade Level reading groups also showed the strongest gains for Ideas & Content and Organization. The Below Grade Level group in particular showed gains that nearly doubled the class average, while the Above Grade Level group showed the least improvement. At this grade level, however, there was no corresponding increase in scores for the lowest language group.

	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
2nd Grade						
<i>Overall</i>	0.41	0.39	0.26	0.23	0.22	-0.02
<i>ELPA</i>						
1s & 2s	0.43	0.40	-0.03	0.07	0.12	0.16
3s	-0.14	0.14	0.28	0.47	0.14	-0.03
4s	0.40	0.40	0.23	0.33	0.27	0.00
5s	0.56	0.56	0.36	-0.03	0.28	-0.11
<i>Reading</i>						
AAGL	0.47	0.42	0.28	0.33	0.35	0.22
SR	0.11	0.39	0.00	0.11	0.30	-0.17
AR	0.56	0.37	0.41	0.20	0.04	-0.18
4th Grade						
<i>Overall</i>	0.28	0.28	0.27	0.18	0.20	-0.06
<i>ELPA</i>						
Monitored	0.25	0.26	0.22	0.07	0.43	0.14
1s & 2s	0.57	0.67	0.83	0.53	0.30	0.20
3s	0.17	0.10	0.03	0.13	-0.12	-0.42
<i>Reading</i>						
AGL	0.00	-0.07	-0.17	-0.07	0.23	-0.07
OL	0.24	0.28	0.33	0.20	0.31	0.24
APP	0.46	0.34	0.46	0.00	-0.04	-0.63
BGL	0.50	0.56	0.46	0.48	0.33	0.02

Table 6.5. Comparison of average change in scores by subgroup, 2nd and 4th grades

Comparison of the average change in Voice scores for second and fourth grade raise some interesting questions about what this trait really means. For both classes, average improvement in the Voice category was about a quarter of a point. In the second grade this meant that improvement for Voice was somewhat weaker than for the two main traits, but in fourth grade, it put Voice on par with Ideas & Content and Organization. The rubrics' descriptions for Voice are nearly identical for both grade levels, but it may nonetheless be the case that the way teachers think of Voice changes from one grade to the other. If what second grade teachers thinks of as a strong "Voice"

involves more interpersonal features, for example, while fourth grade teachers value a more confident presentation of information, it would make sense that gains for this trait were stronger in fourth grade than in second, as the argument unit would reinforce what was already valued in the latter case but not in the former. This would also help explain why Voice tracks more closely with Ideas & Content and Organization in the fourth grade, as the Functional Grammar approach's clear view of the kind of information that needs to be presented and the order in which to do so, as well as its focus on using modality to express certainty, would presumably contribute to a more self-assured authorial voice. Inter-rater agreement percentages for the Voice trait, which in several cases were somewhat lower than other traits on the same tasks, also suggest that it is possible that there is less consensus among teachers themselves regarding what this construct involves.

For those same reasons, Ideas & Content and Organization were the rubric traits expected to be most positively affected by the intervention, and the results of independent scoring suggest that this was the case. The average change in Ideas & Content for the entire second-grade class was a 0.41-point improvement, and Organization improved by an average of 0.39. The fourth-grade class received scores averaging 0.28 higher on both of these traits (followed closely by Voice). In all cases, these numbers reflected stronger improvement on these two traits than on any other, in some cases much stronger. While discussion of averages cannot help but obscure the range of individual performance within each group or subgroup, comparing averages across traits reveals that pre- to post-intervention change is concentrated in the areas that were given most attention in the argument unit.

Some of the differences discussed here can no doubt be accounted for by the design of the 6+1 Traits rubric itself, which even in a version designed to evaluate persuasive writing is still fairly general and makes no mention of the structural components (i.e., stages) of written arguments, providing no focus on the purposes of argumentation and what is expected in the genre. Though beyond the scope of the present study, the question of why Evidence/Reason seemed to be valued less by independent raters than stage analysis could be investigated empirically by having teachers who had participated in FG professional development apply the 6+1 Traits rubric and comparing their scores to those given by teachers who were not participants. Presumably, participants would have been sensitized to the Reason element and their perceptions of the strength of a given argument would differ accordingly (such that an argument with a poorly composed Reason or no Reason at all would be rated lower, most likely on Ideas & Content and/or Organization, and vice versa). Along these same lines, student writing could be scored by trained raters using a primary trait or holistic rubric supported by a more robust concept of argumentation (e.g., Reznitskaya, Kuo, Glina, & Anderson, 2008). Such studies would help clarify the extent to which differences in scoring can be attributed to teacher knowledge and experience versus the instruments themselves, an important distinction for both writing research and classroom implementation.

Evaluative Framework

A major contribution of this research is the elaboration of a framework for the evaluation of argumentative writing. While the 6+1 Traits rubric and others like it provide teachers with an expedient way of judging the overall quality of students'

writing, these instruments lack the specificity that could make them more useful tools. The evaluative framework proposed here provides concrete, stage-specific guidelines to assist teachers in evaluating writing and generating fine-grained, individualized feedback. In addressing elements related to logic, language, and scientific understanding, it also brings together a number of facets of fact-based argumentation that existing rubrics and analytic frameworks do not integrate, be it because they were not sufficiently informed by any one theory of argumentation, because they were designed to evaluate everyday or literature-based arguments (e.g., Harris, Graham, & Mason, 2006; Reznitskaya, Anderson, & Kuo, 2007), or because they emphasize reasoning without adequately conceptualizing the linguistic resources involved in building an argument (e.g., Knudson, 1992a; McCann, 1989; Reznitskaya, Kuo, Glina, & Anderson, 2008; Voss & Means, 1991).

Table 6.6 shows the five dimensions of the framework—Logic, Factual Correctness, Function, Consistency, and Features—and the way they are operationalized at the Evidence/Reason and Counterargument/Evaluation stages. The Logic dimension, the most “holistic” of the five, prompts the rater to consider whether the reasoning behind the stage pair is sound overall, or what, if any, additional information is needed for the reader to understand the student’s thinking. The Factual Correctness dimension is a check on the student’s accurate reporting of the source material that provides the basis for the argument. It is included as a distinct dimension because it is entirely possible to craft a seemingly logical argument based on a misunderstanding of the facts. In pedagogical applications, separating Factual Correctness from the other four dimensions would allow a teacher to see that, for example, a student appears to have control of attributes specific

to argumentation but may need clarification of the facts relevant to the particular issue that is being argued.

Evaluative Dimension	Criteria: Evidence and Reason	Criteria: Counterargument and Evaluation
I. Logic	Does it make sense? Is the reader able to follow the argument without needing additional background?	Does it make sense? Is the reader able to follow the argument without needing additional background?
II. Factual Correctness	Do(es) the statement(s) accurately reflect the source material and/or classroom conversation? Does the writer understand (is the writer able to explain) the concepts?	Do(es) the statement(s) accurately reflect the source material and/or classroom conversation? Does the writer understand (is the writer able to explain) the concepts?
III. Function	Is the general shape of the stages discernible? Do Evidence and Reason “do” what they’re supposed to? Does the author use language to signal movement between stages?	Is the general shape of the stages discernible? Do Counterargument and Evaluation “do” what they’re supposed to? Does the author use language to signal movement between stages?
IV. Consistency	Do Evidence and Reason, in fact, support the writer’s claim? Does the Reason address the particular piece of evidence presented?	Is the Counterargument one that would be plausible for someone with a different claim to make? Is it, in fact, counter to the argument the writer is making? Does the Evaluation effectively address the counterargument presented (do the stages “match”)?
V. Features	Does the writer use language features associated with these stages where appropriate (e.g., modality to make predictions, discuss the likelihood of a particular outcome or how certain scientists are about facts)?	Does the writer use language features associated with these stages where appropriate (e.g., modality to discuss the likelihood of a particular outcome or how certain scientists are about facts, use of contrastive or concessive resources as needed)?

Table 6.6. Evaluative framework for argument (fourth-grade version)

Function, Consistency, and Features are closely related dimensions that draw raters’ attention to the structural and linguistic aspects of the argument. Function and Consistency represent two angles from which to evaluate the functional effectiveness of a stage pair. Function is an approximation of students’ implicit genre knowledge (Kamberelis, 1999). Students who perform well on this dimension show understanding of the function of each stage (i.e., the particular ways it is intended to help the argument unfold) and have written something that, overall, conforms to these expectations. What they have written is easily identifiable as, for example, Evidence followed by a Reason.

Consistency, for its part, serves as a check on the internal relationships between elements of the argument; although it seems counterintuitive, the data analyzed for this study show that students can recreate the general shape of a stage pair appropriately while still needing to improve the quality of the logical relationships that constitute it:

Some people say the bears will be safe at the zoos. I disagree because cubs won't know how to catch fish. (Sample 98)

A teacher using this framework to evaluate Sample 98 could tell the second-grade writer of Sample 98 that he did a good job of writing his Counterargument and starting his Evaluation (Function), and engage him in a conversation about whether he wanted to add more information to his Evaluation to show how it related to the Counterargument or choose a different Evaluation that more directly addressed it (Consistency). Another example of how these dimensions diverge comes from the handful of students in the sample at both grade levels that crafted Evidence and Reason pairs that, while otherwise well written, did not actually support the claim they were making.

While the Consistency dimension is similar for both grades in its emphasis on “matching”, the description for fourth grade also stipulates that the Counterargument be plausible. This is a rather minor, but still important, point. The constraints built into the second-grade source material combined with thorough modeling and discussion made this a nonissue for younger students. Older students, however, were given much more latitude in selecting ideas for inclusion in their arguments. This increased freedom is entirely appropriate as students become more competent writers and thinkers, but as they learn how to argue credibly it must be recognized that part of maintaining that credibility is treating opposing viewpoints fairly. In classroom instruction, students were encouraged

to pick the strongest possible counterargument to respond to, and this aspect of the fourth-grade framework evaluates whether they made a reasonable effort to do so. Ultimately, Consistency in the form of writing an Evaluation that effectively responds to a Counterargument is meaningless if that counterargument is one that would never be made by someone who was seriously arguing.

Finally, the Features dimension provides dedicated space for raters to assess whether students seem to have control over the language points emphasized in instruction. The description of Function also makes reference to language used to signal movement between stages, and in some cases, especially in the more simplified second-grade writing, there will be overlap on this point. The key distinction is one of foregrounding: Function takes language use into consideration as one of several components contributing to the overall shape of the stage, while Features takes language features as its primary focus. Building meaning in context is, of course, the goal of language instruction in the FG curriculum, and this holds true for the Features dimension; when trying to assess how well a student has learned how to use particular language features, the context simply narrows. Raters using this framework should take into account not merely whether a grammatical form was properly constructed, but whether the student's linguistic choices were appropriate given their rhetorical goals. The exact features evaluated with this dimension may vary according to what is emphasized in a given curriculum. In the case of a future version of the FG argument curriculum, modified per the recommendations made in this chapter, teachers would likely want to follow up on their students' use of modality (as relates to making predictions, discussing the likelihood of a particular outcome, or relating how certain scientists are about facts)

for both stage pairs, and contrastive and/or concessive resources for Counterargument/Evaluation. As with other dimensions, analysis of Features is meant to be used as a formative assessment, providing feedback to individual students about their own performance and helping teachers identify points for further review.

The evaluative framework described here is clearly a very different approach to writing assessment than the rubrics frequently used in Oak Grove and districts around the country (Graham & Perin, 2007). Using it in the classroom would mean a substantial investment of teachers' time and effort in first acquiring deep knowledge of argumentation and in actually applying the framework and sharing the results with students. Ultimately, however, this is an investment that is utterly worth making. In an education landscape where teachers consistently report being underprepared to teach writing (Gilbert & Graham, 2010, Read & Landon-Hays, 2013), an instrument that both occasions teacher learning and itself provides concrete guidance for improving student writing has the potential to be a powerful addition to instructional toolkits.

Comparison of Student Writing and the Common Core State Standards

A comparison of the findings of this study's stage analysis with the expectations for argumentation outlined in the Common Core State Standards makes clear that it is possible for students to meet, and in some cases surpass, these goals. The argument ("opinion") writing standards for Grade 2 ask children to:

Write opinion pieces in which they introduce the topic or book they are writing about, state an opinion, supply reasons that support the opinion, use linking words (e.g., *because*, *and*, *also*) to connect opinion and reasons, and provide a

concluding statement or section. (Common Core State Standards Initiative, 2010, p. 19)

Second graders co-constructed their introductions, or Descriptions of Problem, but wrote their own opinions, or Claims, with some scaffolding. The CCSS term “reasons” appears to play a role similar to that of Evidence in the FG curriculum; as discussed above, students were able to select and present evidence in support of their position, and, in providing additional explanation of the relationship between their Claim and Evidence (i.e. in writing their Reasons), went beyond the requirements in the Standards. Their use of linking words and their writing of concluding statements were also consistent with the expectations for second graders. In writing basic Counterarguments and Evaluations, second graders in this study exceeded expectations by several years: counterarguments do not appear as a requirement in the CCSS until Grade 7.

In Grade 4, the CCSS’ expectation is that students:

Write opinion pieces on topics or texts, supporting a point of view with reasons and information [in which they] a. Introduce a topic or text clearly, state an opinion, and create an organizational structure in which related ideas are grouped to support the writer’s purpose; b. Provide reasons that are supported by facts and details; c. Link opinion and reasons using words and phrases (e.g., *for instance, in order to, in addition*). d. Provide a concluding statement or section related to the opinion presented. (Common Core State Standards Initiative, 2010, p. 20)

Fourth graders in this study were, in general, able to meet these expectations, which align in many ways with the FG presentation of argument. Explicit discussion of

the way genre-specific meanings unfold stage by stage provided support for establishing an organizational structure; likewise, fourth-grade writers introduced a topic, stated an opinion, provided reasons related to facts and details (though in this study, reasons were provided in support of facts rather than the other way around), and provided a concluding statement or section. The linking words or phrases suggested by the CCSS were not focused on in this study; rather, the focus on linking phrases in the instructional approach presented here was more specific to argumentation, emphasizing language like “Some people say” to support the specific function played by Counterargument. As in the second grade, Counterargument and Evaluation were themselves evidence of children’s capacity to engage with elements of argumentation beyond those required by the Standards.

Interestingly, the CCSS wording for second and fourth grades includes only a limited emphasis on the strength of written arguments as examined in this study. Fourth graders, for example, are asked to introduce a topic “clearly” and to support reasons with facts but not, for example, to “organize the reasons and evidence *logically*” or “support claim(s) with *logical* reasoning and *relevant* evidence,” (emphasis added) as the Grade 7 standards are written (Common Core State Standards Initiative, 2010, p. 42). This suggests that a higher proportion of students in this study would be seen to have performed successfully if evaluated using the criteria implied by the Standards rather than the criteria for stage analysis, which not only accounted for the presence of these elements but strongly emphasized the effectiveness of each. In addition, it is only in Grade 6 that the Standards begin to ask students to “establish and maintain a formal style” (Common Core State Standards Initiative, 2010, p. 42) in their arguments. While

the authors of the Standards do not elaborate on what they consider to be a formal style, it is likely that this construct involves a move towards a less personal, more measured and authoritative tone such as the one modeled in the FG approach. If this is the case, the study participants' use of limited personal pronouns and carefully chosen modal meanings indicate that formal style is another construct with which children are able to engage sooner than the Standards would ask them to. In sum, then, these findings are encouraging with respect to children's capacity to meet the goals set out for them by the Common Core; they also raise interesting questions about whether, for some elements of argumentation, the bar can or should be set even higher.

Limitations and Directions for Future Investigation

Several issues limit the conclusions that can be drawn from this study's findings. This section discusses limitations related to sampling and pedagogical considerations and their implications for future research.

Sampling. There are a number of sampling considerations that should be addressed through additional research. The relatively small sample, consisting of just one class at each grade level under study, limits generalizability; using this data it is not possible to conclude whether the performance of these students is representative of other students of the same age, or whether it was a result of conditions or characteristics specific to these classes or the Oak Grove district. Similarly, the composition of the samples leaves some important questions unanswered. The participants in this study were all currently classified as ELLs, or had been in the recent past, and all of them were exposed to a form of Arabic as a first or additional language; there were no students in the sample that were exclusive, native English speakers. Whether there would be a wider

range in performance in classrooms with more non-ELLs, or with a wider variety of languages spoken, remains to be seen. These latter issues in particular will need to be addressed in a more demographically diverse district. Ideally, this research would involve expanding the design to include additional grade levels and would take place in schools in which multiple classes at each level could be studied, thereby facilitating future experimental studies and enabling a longitudinal design that could show differences between grade levels.

Pedagogical considerations. Future research should also address some limitations of the pedagogical design. The instruction described in this study served as a pilot for the argument unit developed for a larger study. The research team approached instruction as an opportunity to refine the argument lessons, and recent iterations of that curriculum have already incorporated some of what was learned during implementation regarding how to sequence instruction and how best to explain certain concepts (e.g., see description of Counterargument in fourth grade above). The pedagogical recommendations developed in this and previous chapters are potential contributions to enhance future instruction.

Another important factor that shaped pedagogy and, accordingly, the results of this study was the amount of time allotted for the unit. Where the majority of instructional interventions described in Chapter 2 took place over the space of several weeks or months, scheduling constraints meant that this study's argument lessons took place over five to six school days. This may well have had positive consequences for retention of material from one lesson to the next, but it also meant that finite limits were placed on the amount of practice students had with deconstructing and analyzing sample

arguments and writing their own. Future research designs should aim for an expanded timeline that would allow for reinforcement of key concepts, as well as follow-up studies measuring student performance after additional time has passed.

In addition, there is a need to study teacher-led implementation of argument pedagogy. Researcher-led implementation was appropriate for piloting and refining FG argument lessons, and the research team members' ability to draw on theoretical knowledge helped strengthen the presentation of these concepts in the curriculum. Observing implementation in applied contexts ("Doing research on teacher implementation in classroom contexts"), however, will be necessary in order to learn which modifications may be needed to prepare the curriculum for practitioner use, including the development of support materials for teachers who may not have sufficient background knowledge of argumentation and/or FG constructs.

Finally, the treatment of issues of audience, purpose, and authenticity should be given special consideration in future iterations of this argument curriculum and others like it. These issues have been found to affect students' writing and reading performance (see, e.g., Block, 2013; Graham, McKeown, Kiuahara, & Harris, 2012; Midgette, Haria, & MacArthur, 2008). Attention was paid to the overall social purpose of argumentation, and especially the functional purposes of its stages, throughout the argument unit. However, while some attempt was made to have students consider an audience for their arguments (e.g., classmates on the other side of the issue in second grade, the Australian government or students in the class across the hall in fourth grade), our approach to doing so was not systematic, and the absence of a clearly-defined audience at whom the students could direct their thinking and writing potentially weakened focus on social purpose.

As described in Chapter 3, a primary concern in selecting topics and creating source texts was keeping the scientific content learned alongside the genre content cognitively manageable. To that end, the grizzly bears and cane toads issues, while not germane to Oak Grove students' everyday lives, built strategically on the concept of the consequences of human interference on various ecosystems; the grizzly bears issue also drew directly on the second graders' previous study of the Alaskan River ecosystem and also prepared them for a field trip to the local zoo. This consideration, coupled with the fact that the FG approach intentionally focuses on school genres, reduced what a number of scholars would consider the authenticity of the reading and writing activities, which also may have reduced the instruction's effectiveness. Purcell-Gates, Duke, and Martineau (1997), for example, found that degree of authenticity—defined as “reading and writing of...genres that occur *outside* of a learning-to-read-and-write context and purpose, and...reading and writing those texts for the purposes for which they are read or written outside of a learning-to-read-and-write context and purpose” (p. 14)—had a strong relationship to second and third graders' growth along a number of dimensions, especially when coupled with explicit instruction in the language features characteristic of the targeted scientific text types (conversely, explicit instruction coupled with a low degree of authenticity had much weaker effects on student growth). In this view the goal of teaching a form of argumentation valued in schooling contexts would be by definition a low-authenticity activity. Taking a different view of authenticity, it could be argued that given the amount of time students spend in school weekly and over the years, school genres are themselves highly authentic—certainly no less than professional genres produced by adults in work settings. The FG focus on school genres also emphasizes

language features associated with academic registers, and a shift in audience may also produce a shift in register that is inconsistent with instructional goals related to those features. Still, it may be possible to further enhance students' learning by providing them with a higher-authenticity *purpose* for producing this school genre, including establishing a concrete and relevant audience, and future research should aim to do so.

Testing and elaboration of evaluative framework. The evaluative framework presented in this study is subject to the same sampling and pedagogical considerations mentioned above. The framework was elaborated using texts gathered in just one classroom at each grade level, and should be applied in a variety of contexts to determine whether the descriptors are comprehensive or if they need to be further refined. In addition, testing by teachers will help determine which modifications may be necessary in order to facilitate integration into existing pedagogical contexts. This may involve producing supplementary materials for an administrator's guide, including suggestions for reconciling the foci of the framework with elementary teachers' need to monitor students' use of surface-level features. Given its potential use as an analytic tool for researchers and practitioners alike, additional study of the framework will be incorporated into future research.

Concluding Thoughts

Analysis of second and fourth graders' writing showed that students in this study produced arguments that were more complete and more effective than children their age are frequently given credit for. In a pedagogical context that provided rich support for language and content, students across academic and language proficiency levels were able to take positions on issues removed from their everyday lives; present and interpret

factual evidence in support of those positions; and identify and respond to counterarguments. They used linguistic tools appropriate for conveying a reasoned academic examination of the issues they wrote about. The argumentative writing produced prior to students' participation in the argument unit did not share these defining features. Rather, students wrote strong academic arguments as a result of instruction that built a meaningful context for explicit discussion of the genre's function and form. Through a recursive cycle of deconstruction, joint construction, and independent construction, students explored an argument genre at a level that went far beyond simple description or listing of its features. They identified how sample texts unfolded and how each step helped their authors achieve their purposes; examined the role key language features played in conveying meaning; enacted argumentation through discussion and debate with their peers; and built their own arguments stage by stage, reflecting on their thinking and writing in whole-group and small-group conversations as they went. In doing so, students not only met but exceeded many of the expectations established for their respective grade levels in the Common Core State Standards, a finding suggesting that with sufficient instructional support in place, achieving the challenging new goals set out by the CCSS is indeed within students' reach.

An unexpected finding of this study was the lack of relationship between students' academic achievement, as measured by reading level, or language proficiency, as measured by a statewide standardized test, and their performance on different dimensions of the argumentative writing task. The effectiveness of students' writing at one stage versus the next was independent of subgroup membership. This is an indication that the pedagogy supporting this writing might be especially effective in buoying the

performance of the students who are normally weakest, and who under normal circumstances would be expected to perform less well as a group than their higher-achieving peers. Given U.S. students' ongoing struggles with writing performance, especially as regards persuasive/argumentative writing (NAEP, 1994, 1999; Perie, Grigg, & Donahue, 2005), and their teachers' search for instructionally sound approaches to addressing the problem (Gilbert & Graham, 2010; Newell, Beach, Smith, & VanDerHeide, 2011), these results provide cause for optimism.

The fact that most students were able to execute the argumentative writing task with some level of success, however, does not mean that there is no room for improvement. Writing analysis showed that students' performance was not uniform across stages. Effective writing in one part of an argument did not mean that other pieces would be similarly effective; nor, for that matter, did weak writing on one stage mean weak writing across the board. This kind of nuance is not captured by the rubrics typically used to evaluate student writing in classrooms or on standardized tests; nor, for the most part, is it reflected in the measures used by those studying children's argumentation. A key strength of the analytic approach used in this study is that it makes possible a detailed accounting of the strengths and weaknesses of individual arguments and in turn allows for characterization of the range of performance within a given group, including identification of stages of argumentation that are especially in need of more robust instruction. Analysis also revealed the influence that informational source texts can have on the quality of children's written arguments, suggesting the importance of carefully considering the role of instructional features beyond those directly pertaining to teaching about and supporting the writing itself.

The kind of empirical evidence provided by this study regarding the structural, logical, and linguistic features of argumentation with which students are likely to need additional support not only suggests promising avenues for future research but also provides clear guidance for how to improve instruction, guidance which is sorely needed when teachers are so often left to rely on their best guesses in crafting writing pedagogy to address newer, more challenging standards. The recommendations resulting from this study, coupled with the detailed evaluative framework, are concrete steps toward helping teachers know what to focus on in supporting the development of their young students' written arguments, and to know that with that support, their students can produce more successful writing than we thought possible.

Appendix A: Focal student interview protocol

1. Will you read your argument out loud?
2. What was your claim?
3. Are there any words in your claim that tell me how sure you are or how strongly you feel about the recommendation you're making?
4. What evidence did you use?
5. What reason did you use to explain why it was important?
6. Are there any words here that tell me how sure you are about your reason?
7. Which counterargument did you choose?
8. How did you start off the counterargument?
9. What were the words you used to begin?
10. What was your evaluation of that counterargument?
11. What are the words that start that?
12. What was your goal for the argument? If someone read what you wrote how would you want them to react? What would you want them to think?
13. Do you think what you wrote would convince them? Why?
14. Is there anything else you want to tell me about what you wrote or anything else you'd like to learn about?
15. What was your favorite thing we did this week?
16. What was the hardest thing about what we did this week?

Appendix B: Scoring materials for independent raters

	6	5	4	3	2	1
IDEAS & CONTENT	This paper gives a strong and clearly stated and supported opinion. Provides reasons-- supported with facts and details. Maintains a clear, consistent stand from start to finish. Provides a compelling concluding statement.	This paper gives a clearly stated and supported opinion. Provides reasons that support the opinion. Maintains a clear, consistent stand from start to finish. Provides a concluding statement related to opinion.	This paper gives a somewhat clearly stated supported opinion. Opinion is evident. Provides reasons that support the opinion.	Opinion statement is evident. Opinion is written in a basic sentence. Opinion is noted. One reason is given to support opinion.	The writing is only occasionally clear and focused. Ideas and content are underdeveloped. Lacks topic sentence. Opinion statement is not evident.	As yet, the paper has no sense of purpose. No opinion stated.
ORGANIZATION	The writer's control over organization effectively moves the reader through the paper. The introduction is thoughtful and effective. There is a clear beginning, middle, and end. The body provides reasons with facts and details that support the opinion. A concluding statement is evident.	The writer's control over organization effectively moves the reader through the paper. The introduction is thoughtful and effective. There is a clear beginning, middle, and end. The body provides reasons that support the opinion. A concluding statement is evident.	The response is generally coherent, and its organization is functional. The introduction states a strong opinion. The body provides a reason for the opinion. Provides a concluding statement.	The organizational structure is strong enough to move the reader through the paper without too much confusion. Paper contains a beginning. There is a body of the paper. No conclusion is evident.	There may be little evidence of organizational structure.	The writing lacks a clear sense of order.
VOICE	The voice is extremely confident and convincing and holds the readers attention throughout. The writer is mindful of the audience and connects with the reader.	The writer is mindful of the audience and connects with the reader.	The writer begins to show how he or she really thinks or feels about the topic.	The voice is somewhat convincing and holds the readers attention most of the time. There is a moment of audience awareness, but then fades.	The paper connects with the reader in the most general way.	The reader is not sure why the writer chose this idea for writing.
WORD CHOICE	Words send message in a clear, interesting, and natural way. The words are powerful and engaging . The words paint a clear picture in the reader's mind of the writer's opinion. Writer uses transitions and phrases consistently and accurately.	Words send message in a clear, interesting, and natural way. Uses transition words, phrases, or clauses such as: consequently, in order to, etc. to connect to the opinion and reasons.	Begins to use transition words such as: because, and...to connect opinion and reasons. The reader begins to understand the writer's opinion.	Language is functional, even if it lacks much energy. It is easy to figure out the writer's meaning on a general level. Vocabulary may be basic. Run-on sentences exist.	Few words can be identified.	Few, if any, recognizable words are present.
SENTENCE FLUENCY	Sentences are skillfully written and original. Reading aloud is a breeze and keeps the reader's interest.	Sentences show variety and are easy to read and understand. Invites reading with enthusiasm.	The sentences are varied but should flow more smoothly. Sentences start differently and are of varied length. Can read aloud with expression.	A better variety of sentences is needed. Sentences do not read smoothly. Piece can be read but construction is awkward. Long sentences go on and on.	Simple sentences. Sounds choppy when read.	Few sentences are written well. Cannot read aloud.
CONVENTIONS	Demonstrates a good grasp of writing conventions and uses conventions effectively to enhance readability.	Language is well controlled, and occasional lapses in writing conventions are hardly noticeable. Appropriate punctuation and conventions applied.	Lapses in writing conventions are not distracting. Spelling is correct and editing required.	Reasonable control over a limited range of standard writing conventions. At other times, errors are distracting. Spelling is inconsistent.	Limited control over writing conventions may make the writing difficult to understand.	Errors in spelling, punctuation, capitalization, grammar and/or paragraphing repeatedly distract the reader and make paper difficult to read.

B1. Oak Grove opinion writing rubric for grades 1-3

	6	5	4	3	2	1
IDEAS & CONTENT	This paper gives a strong and clearly stated supported opinion. Contains intriguing and relevant specific facts, details, and examples to support the opinion. Maintains a clear, consistent stand from start to finish. Provides a compelling concluding statement.	This paper gives a clearly stated supported opinion. Contains relevant specific facts, details, and examples to support the opinion. Maintains a clear, consistent stand from start to finish. Provides a conclusion statement related to opinion	This paper gives a somewhat clearly stated supported opinion. Contains some relevant specific facts, details, and examples to support the opinion. Provides a concluding statement.	Opinion statement is evident. The writer is beginning to define the topic even though development is still basic and general.	The writing is only occasionally clear and focused. Ideas and content are underdeveloped. Lacks topic sentence. Opinion statement is not evident.	As yet, the paper has no sense of purpose. To pull meaning from the text, the reader must make inferences based on sketchy or missing details. Limited information. Does not state an opinion.
ORGANIZATION	The order, or presentation of information is compelling and moves the reader through the paper between ideas.	The writer's control over organization effectively moves the reader through the paper. Clearly states opinion. Provides logically ordered reasons supported by facts and details. Provides a concluding statement that supports the opinion.	The response is generally coherent, and its organization is functional. The statement of opinion is clear. The body provides reasons that are supported by facts and details. A summary is noted.	The organizational structure is strong enough to move the reader through the paper without too much confusion. The statement of opinion is given in the introduction. Some simple supporting details noted in the body. No concluding statement.	There may be little evidence of organizational structure.	The writing lacks a clear sense of direction. Ideas or details seem strung together in a loose or random fashion; there is no identifiable structure.
VOICE	The voice is extremely confident and convincing and holds the readers attention throughout.	The voice is generally confident and convincing and holds the readers attention throughout.	The voice is somewhat confident and convincing and holds the readers attention most of the time.	The voice is somewhat convincing and holds the readers attention most of the time.	The voice is at times convincing and holds the readers attention in parts.	The voice is unprofessional and not appropriate for the purpose and audience.
WORD CHOICE	Words send message in a clear, interesting, and powerful and engaging . Transition words link opinions and reasons. These words signal support and conclusion: such as, finally, in conclusion. Sentences are skillfully written and original. Reading aloud is a breeze and keeps the reader's interest.	Words send message in a clear, interesting, and natural way. Uses transition words, phrases, or clauses such as: consequently, in order to, etc. to connect to the reasons. Sentences show variety and are easy to read and understand. Invites reading with enthusiasm. Language is well controlled, and occasional lapses in writing conventions are hardly noticeable.	Words send message in an effective way, such as the use of linking words that connect opinion and reason (therefore, since, etc.). The sentences are varied but should flow more smoothly. Can read aloud with expression. Lapses in writing conventions are not distracting.	Language is functional, even if it lacks much energy. It is easy to figure out the writer's meaning on a general level. Vocabulary may be basic.	Language is functional but it is difficult to figure out the writer's meaning on a general level. Vocabulary maybe limited.	Demonstrates a limited vocabulary or has not searched for words to convey specific meaning.
SENTENCE FLUENCY	Sentences are skillfully written and original. Reading aloud is a breeze and keeps the reader's interest.	Sentences show variety and are easy to read and understand. Invites reading with enthusiasm. Language is well controlled, and occasional lapses in writing conventions are hardly noticeable.	The sentences are varied but should flow more smoothly. Can read aloud with expression. Lapses in writing conventions are not distracting.	A better variety of sentences is needed. Sentences do not read smoothly.	Simple sentences. Sounds choppy when read.	Few sentences are written well. Cannot read aloud.
CONVENTIONS	Demonstrates a good grasp of writing conventions and uses conventions effectively to enhance readability.	Language is well controlled, and occasional lapses in writing conventions are hardly noticeable.	Lapses in writing conventions are not distracting.	Reasonable control over a limited range of standard writing conventions. At other times, errors are distracting.	Limited control over writing conventions may make the writing difficult to understand.	Errors in spelling, punctuation, capitalization, grammar and/or paragraphing repeatedly distract the reader and make paper difficult to read.

B2. Oak Grove opinion writing rubric for grades 4-5

Excerpted from rater instructions:

Task One (Grade 2):

Prompt: Should people put grizzly bears in zoos? Use specific evidence and reasons to explain or why not.

Description: Students read and discussed texts about endangered species and the pros and cons of using zoos to save them. Their task was to take a stand on whether grizzly bears should be put in zoos, using facts about grizzly bears to support their reasoning. Important facts include that grizzly bears need lots of space in the wild (they have a range the size of 20 Dearborns); that grizzly bear cubs stay with their mothers for 2-3 years and learn to hunt and fish during this time; and that grizzlies' food sources, especially salmon, may be limited in the wild.

Task One (Grade 4):

Prompt: Scientists in Australia think using meat ants might be one effective way to stop cane toads from spreading and hurting more organisms. What do you think of this option? Is it a good option Australia should act on now, a good option they should study more, or not a good option? Why?

Description: Students read and discussed texts about cane toads. Cane toads were brought to Australia to eat beetles that were destroying sugar cane crops in the early 1900s, but turned out to be highly destructive to the local ecosystem because they reproduce rapidly, eat large amounts of native insects and other species, and produce an extremely potent toxin that quickly kills almost any potential predator. Students were presented with information about meat ants, an aggressive ant species that sometimes eats animal flesh, and asked to evaluate them as a potential solution (see prompt). Important information they may have taken into account as they wrote includes:

- Meat ants are not affected by cane toads' poison
- Native toads flee from meat ants, but cane toads do not recognize them as a predator and sit still while attacked
- Despite their name, meat ants' diet normally does not consist primarily of meat, but of honeydew from caterpillars. In their natural habitat, this is a mutually beneficial relationship, as the ants provide protection for the caterpillars.
- Preliminary studies suggest that meat ants would be a safe solution, but this theory has not been tested on a larger scale in the wild.

Task Two (Both Grades):

Prompt: In Florida, there are many hotels along the sandy beach. Every morning, the people who work at these hotels go out and remove the beach wrack so the hotel guests can swim on a clean beach. Should the hotel workers take away the beach wrack? Use specific evidence and reasons to explain why or why not.

Description: Students read and discussed a text about the sandy beach habitat. One part of the sandy beach habitat is *beach wrack*. Beach wrack is piles of seaweed, shells, trash, and dead animals. It provides a home for organisms including kelp flies, pseudoscorpions, rove beetles, and beach hoppers. Organisms from other parts of the sandy beach habitat, like birds and crabs, depend on the beach wrack organisms for food. After learning about the sandy beach habitat, students were presented with this prompt. They *did not* read any additional material about hotels removing beach wrack from the shore.

B3. Background information provided to independent raters

Sample Number	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions	Comments
<i>Example</i>	4	3	5	4	3	2	
Grizzly Bears							
2							
18							
31							
128							
153							
156							
Sandy Beach							
3							
7							
21							
87							
89							
92							
97							
133							
134							
144							
147							
	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions	
Comments about these samples as a whole (for example, what stood out? How does this compare to other second graders' persuasive writing?):							

B4. Sample blank scoring sheet (second grade)

Appendix C: Graphic organizers for planning and writing, second grade

What do you think? Take a stand

Describe the problem: Why did people want to put bald eagles in zoos?

Make a claim: Give your opinion. Are zoos a good solution or a bad solution?

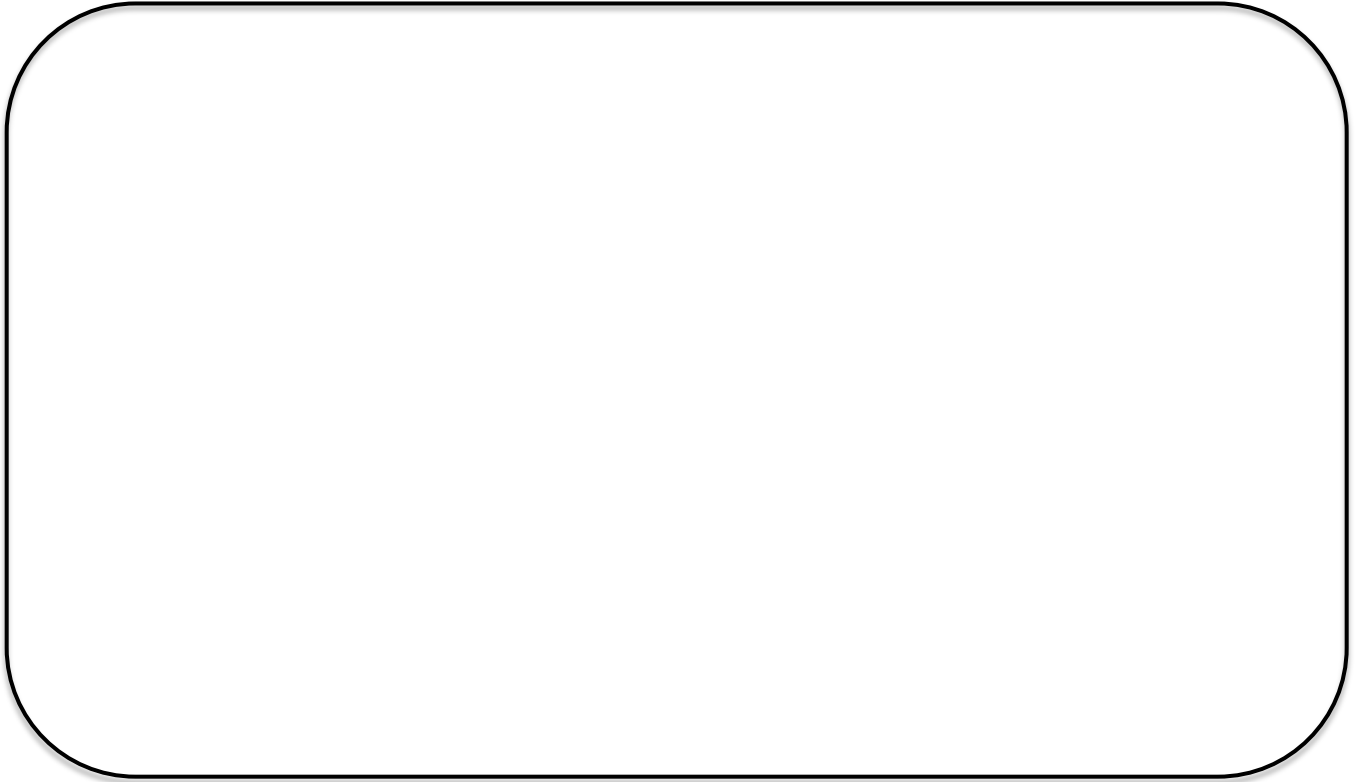
Give evidence and a reason: What information from the text supports your claim? *How* does that evidence support your claim?

State a counterargument and evaluate it: Share information that the other side might use as evidence. Why do you disagree?

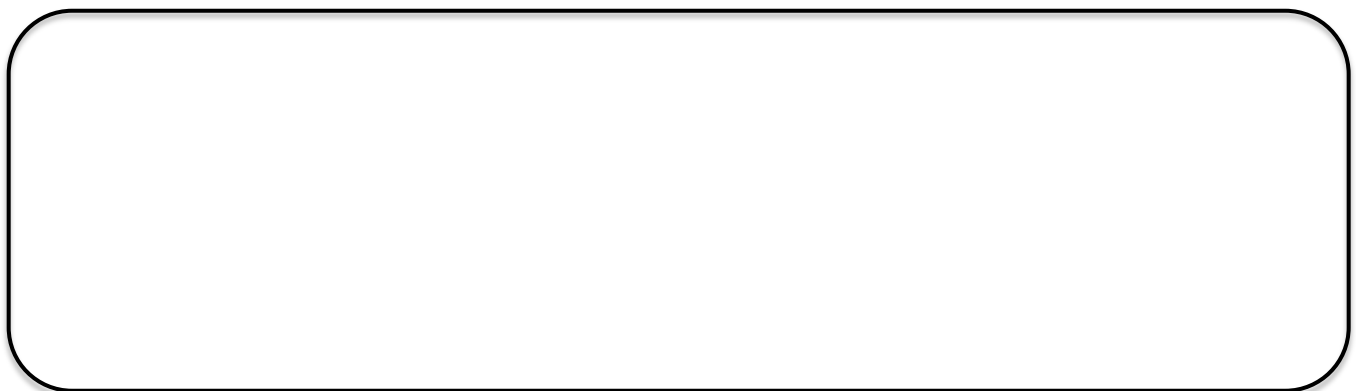
C1. Text for argument poster

Name: _____

Description of Problem: Give your reader important background information on the grizzly bear problem. Why do people want to put grizzlies in zoos?



Claim: Give your opinion. Are zoos a good solution or a bad one?



C2. Description of Problem and Claim scaffolds

Name: _____

Evidence: A piece of information from the text that supports your claim.

Reason: How or why does this evidence help prove your claim? How does this information relate to solving the grizzly bear problem?

C3. Evidence and Reason scaffold

Name: _____

Counterargument: Share a point that someone presenting the other side of the argument might use for evidence.

Evaluation of counterargument: Why do you disagree with the counterargument? Why doesn't it change your mind?

Appendix D: Graphic organizers for planning and writing, fourth grade

Name: _____

Description of Problem: Give your reader important background information on the cane toad problem.



Claim: Briefly introduce the proposed solution you'll be writing about and state your position on that issue.



D1. Description of Problem and Claim scaffolds

Name: _____

Circle your claim (your position on the issue) so far:

Good, act now!

Good, study more

Bad, don't use

Evidence: A piece of information from the text that supports your claim.

Reason: How or why does this evidence support your claim? How and why does the evidence help or hurt the cane toad problem?

D2. Evidence and Reason scaffold

Prompt: Scientists in Australia think using meat ants might be one effective way to stop cane toads from spreading and hurting more organisms. What do you think of this idea? Is it: a) a good option Australia should act on now, b) a good option that Australia should study more before using, or c) a bad option that they shouldn't use? Why?			
Evidence: A piece of information from the text that supports...	Good, Act now	Good, Study more	Bad, Don't use

D3. Evidence collection sheet

Name: _____

Counterargument: Share a point that someone presenting another side of the argument might use for evidence.

✍ Notes for evaluating the counterargument: ✍

What part of the counterargument do you agree with? Why?	On a scale of 1-10, how strongly do you agree ?
--	--

What part of the counterargument do you disagree with? Why?	On a scale of 1-10, how strongly do you disagree ?
---	---

Evaluation of counterargument: Use your notes to respond to the counterargument. Use language that shows how strongly you agree or disagree, and explain why.

D4. Counterargument scaffold

Organizing Your Argument

Describe the **ISSUE**: Give the reader important background information

State your **CLAIM**: What's your position on this issue?

Provide **EVIDENCE & the REASON** it supports your claim.

Provide more **Evidence & Reasons** supporting your claim if possible or necessary

Provide a **COUNTERARGUMENT** and **EVALUATE** it.

Briefly re-state your **CLAIM**. Why is the issue and your position important?

D5. Argument organization overview

Appendix E: Bald Eagles text

Saving Bald Eagles



Bald eagles are the national bird of the United States. They live in many places across the country. They are an important part of the Alaskan River ecosystem.

Bald eagles need a large area to live in. Their wings can be up to eight feet across. They like to live in tall trees, near water and far away from humans. Bald eagles don't live in groups. They spend most of their lives with one other eagle. Bald eagles are carnivores, and they eat fish, including salmon.

Bald eagles used to be an *endangered species*. This means that there were very few eagles left. Some eagles lost their homes because people cut down trees to build cities. Sometimes the water where bald eagles fished became poisoned with chemicals. The eagles ate poisoned fish, and then the eggs they laid weren't healthy. Not many new baby eagles were born.

When people realized that bald eagles were endangered, they tried to save them. For example, people stopped using the chemical that was poisoning the eagles. Some bald eagles were put in zoos, where they ate clean fish and had healthy babies. They were safe in the zoos, and scientists and visitors to the zoos learned more about bald eagles. Many people started to care about helping them.



Now bald eagles are not endangered. There are many healthy bald eagles in the wild.

Appendix F: Model arguments for second grade

Put Bald Eagles in Zoos!

Bald eagles are an endangered species. There are hardly any bald eagles left. We need to do something soon to protect them. We could put them in zoos to save them.

When bald eagles eat poisoned fish, they don't lay healthy eggs. In zoos, bald eagles won't have to eat fish with chemicals. If eagles eat healthy fish in the zoo, new baby eagles will be born.

Some people say that zoos are a bad idea because bald eagles are used to having a lot of space to fly around. I disagree because zoos can build big cages.

We can save the bald eagles by putting them in zoos.

Don't Put Bald Eagles in Zoos!

Bald eagles are in trouble because they are losing their homes and eating poisoned fish. But we shouldn't put them in zoos.

Zoos are a bad idea because bald eagles like to live far away from humans. If eagles are in zoos, they will have to be around humans all the time. They might be unhappy.

People who think zoos are a good idea say that the bald eagles will be safe in the zoos. But zoos only have space for a few eagles, so lots of wild eagles would still be in danger.

We need to save bald eagles, but we should not use zoos to do it.

Appendix G: Grizzly Bear fact sheet

Grizzly Bear Fact Sheet



- Adult grizzly bears can be up to 7 feet tall. They usually weigh from 200 to 600 pounds. The biggest grizzlies weigh more than 800 pounds!
- Grizzly bears need lots of space. The places they live include mountains, forests, and grassy fields. Each grizzly bear needs a space that is twenty times the size of Dearborn!
- Meat like salmon is only a very small part what grizzlies eat. Almost all of their food comes from plants, berries, wild fruits, nuts, and insects.
- Grizzly bears are at the top of the food chain. This means they don't have any natural predators.
- Grizzly bears usually like to be alone. But baby bears, called *cubs*, stay with their mothers for 2 to 3 years.
- Grizzly bears live in many parts of the world. The map shows where grizzlies live. There used to be around 50,000 grizzly bears in North America. Today, there are only about 1,000 grizzly bears left in the United States, not including Alaska.
- Grizzlies could become endangered soon. Their habitat is shrinking as humans take over more land. Many grizzlies die after being shot by humans. Some people shoot the grizzlies because they are scared of them, but other people hunt them on purpose even though it's illegal.



Appendix H: Cane Toads text

Stopping the Cane Toad Invasion



In the 1930s, farmers in northern Australia had a big problem: beetles were eating the sugar cane crop. Sugar cane is a plant that sugar is made from. It is an important crop for Australia. To stop beetles from destroying the sugar cane, scientists brought the South American cane toad to Australia. They thought the toads would eat the beetles. But the plan didn't work.

Instead, the cane toads turned into an even bigger problem. They didn't eat many beetles, but they ate lots of other things, like bird eggs and smaller frogs. The huge toads also laid lots and lots of eggs in ponds and creeks near the cane fields. Female toads can lay 35,000 eggs at a time! The toads left the cane fields, and soon, they were all over northern Australia. They became serious predators of several kinds of Australian insects. Each toad can eat about two and a half times its body weight in one day—that's a lot of food! Even worse, nothing could stop them because cane toads are very poisonous when eaten. Predators like lizards, snakes, crocodiles, and even household pets began to die because they were poisoned when they tried to eat the toads. Scientists have determined that at least 75 species of turtles and crocodiles are at risk because they don't know that eating the toads will kill them. In one study, about 90% of the big lizards in one area had died because they tried eating the toads. A scientist studying the problem said, "There are all sorts of effects. You take out 90% of the big predators and that really changes the system."

The Australian government has tried to get rid of the cane toads in many ways. So far, though, they have not been able to come up with a good solution. Some potential

solutions can't keep up with the number of new toads that are born each year. Other solutions would put native toads and frogs in danger. Scientists are still hard at work trying to come up with a way to stop the cane toad invasion.

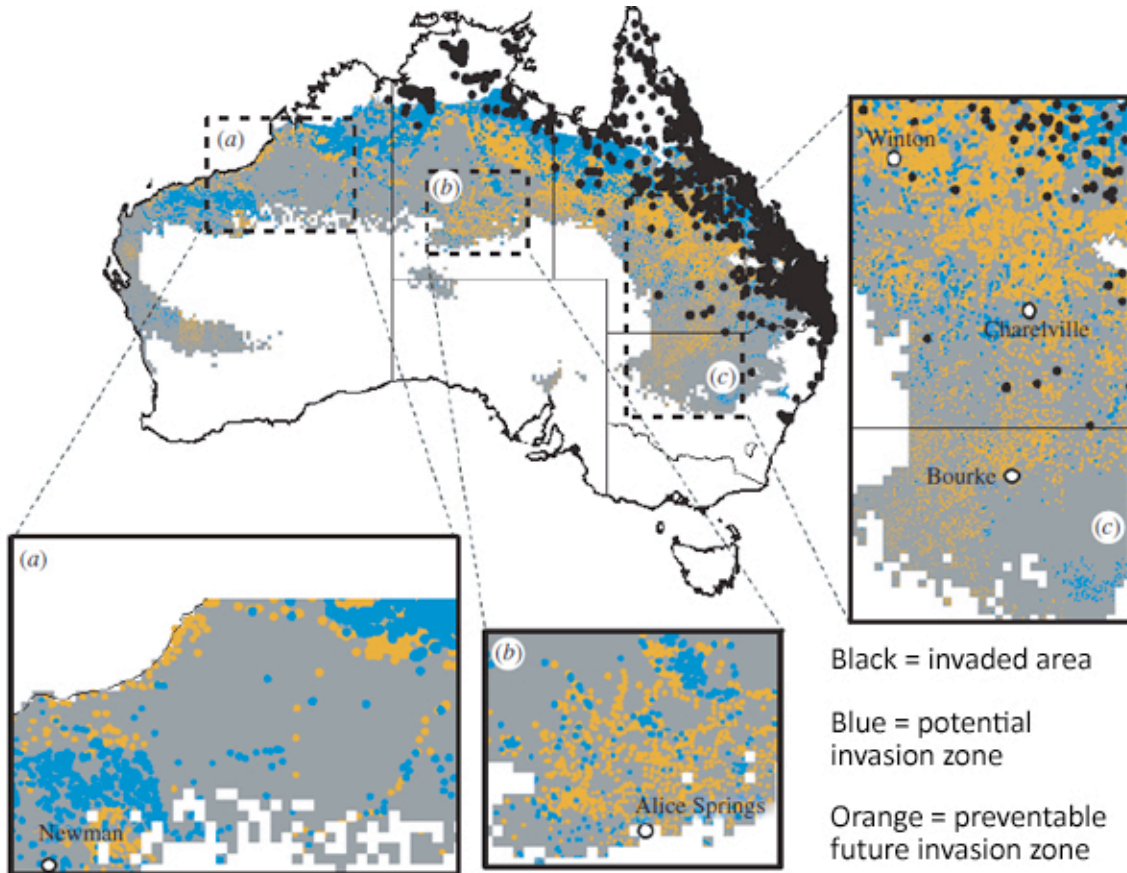


Figure 1. *If cane toads are left alone, scientists predict that they will invade even more land in northern Australia. In the 1930s, just 102 toads were brought to Australia. By 2011, the number of cane toads in the country was estimated to be as many as 2,000,000,000 (two billion)!*

Appendix I: Lungworms text

Using Parasites to Control the Cane Toad Invasion

Scientists at the University of Sydney have identified a potential solution to Australia's cane toad problem: lungworms. Lungworms are parasitic worms that live in toads' lungs. Toads get lungworms when the tiny worms find the toads and crawl into their eyes, or by eating other toads that are already infected. A team of researchers has discovered that the lungworm parasite can kill about 30% of infected baby toads. Young toads that don't die from the parasites tend to stay much smaller, never growing to be as big as uninfected toads. Adult toads that get infected can't move as quickly. Spreading the parasite around so more toads get it might be a good way to control their numbers and keep them from invading new territory.



Parasites often evolve to target particular species. It was originally thought that cane toads left their parasites behind when they came to Australia and that the lungworm parasite came from Australian frogs. This meant that lungworms could not be used as a solution, because they would harm both cane toads and native Australian frogs. However, the University of Sydney scientists have used DNA testing to prove that the lungworms currently found in cane toads actually came from South America. In other words, although they look very similar, cane toad lungworms are not the same species as the lungworms found in Australian frogs.

The researchers are running tests with cane toad lungworms before trying to use them as a solution. They are injecting Australian frogs with the worms to see if this would be very dangerous for them. So far, it looks like cane toad lungworms can infect Australian frogs, but the results are not as serious. The worms tend to get lost inside the frogs and die before finding the lungs. If the worms do make it to the frogs' lungs, the frogs just don't get as sick as toads do. The results of these laboratory tests are promising, but of course, the solution has never been tested in the wild.

Appendix J: Model Arguments for fourth grade

Yes, use the lungworm now!

Cane toads are a huge problem in Australia. 102 toads were brought to the country in the 1930s, but now there are as many as 2,000,000,000. Cane toads breed a lot, eat a lot, and kill a lot of predators with their poison. Scientists think giving the toads lungworms will help this problem. It is worth trying.

One reason lungworms are a good solution is that about 30% of baby toads that get lungworms die. If scientists spread more lungworms in the wild, 30% fewer cane toads would live to be adults. That would slow the invasion down.

Another reason lungworms would be a good way to deal with cane toads is because younger toads that don't die from lungworms still never grow up to be as big as other toads. Healthy cane toads can weigh up to four pounds and eat two and a half times their weight in food. Smaller toads would eat less food and have less of an effect on the insect species they prey on.

Some people might say that spreading lungworms around is a bad idea because we have never tested this idea in nature. So, we can't know exactly what would happen. However, scientists are being careful to test the lungworms on native frog species. The results show that the effects probably won't be that bad. It's worth it for some frogs to get a little sick if it will keep many other species from dying.

If scientists don't act quickly, the cane toad population will spread all the way across Australia. They should try to use lungworms to stop the problem now.

No, don't use the lungworm!

The cane toad invasion has been a disaster for the ecosystems in Australia. Cane toads were brought to Australia to solve a problem, but they have become an even bigger problem themselves. The huge toads eat anything that will fit in their mouths, and nothing can eat them because they are highly poisonous. Scientists think that spreading the cane toad lungworm around could help solve the problem, but it is too dangerous. They should not use it.

First, using lungworms is a bad solution because it has never been tested in nature. When people brought the cane toad to Australia, they never thought it would cause such a big problem, and the same thing could happen with lungworms.

Another reason this is a bad solution is that the worms only kill about 30% of baby toads. That means 70% of toads would still survive. The impact is too small to place the ecosystem in such risk.

One reason to use lungworms is that they do make older toads smaller and slower. But smaller toads still eat a lot of food, and slower toads are still poisonous. In the end, it is highly unlikely that lungworms would make a real difference.

Australia needs a good solution to the cane toad problem, but using lungworm parasites is not the right one. Scientists should look for something better and safer.

Appendix K: Text type overview handout

Argument Text Type

Overall purpose: To take a position on an issue and explain the reason(s) why.

Stages and their purposes:

<i>Description of problem</i>	Familiarizes the reader with important background on the issue or problem.
<i>Claim</i>	A statement that introduces the specific issue to be argued and the writer's position on that issue.
<i>Evidence</i>	A piece of information from the text that supports the overall claim.
<i>Reason</i>	A statement about why/how the evidence supports the claim and interprets the evidence in relation to the larger issue.
<i>Counterargument</i>	A piece of information that someone presenting another side of the argument might use for evidence.
<i>Evaluation of Counterargument</i>	Evaluates the strength of the counterargument and explains what this information means for the writer's position on the issue.
<i>Restatement of claim</i>	Summarizes and restates the writer's position on the issue.

Appendix L: Meat Ants text

Meat Ants

Meat ants, also known as gravel ants, can be found throughout Australia. They originally got their name from being used by farmers to clean carcasses. They live in underground nests of over 64,000 ants and prefer sunny, damp areas.



Meat ants eat during the day. Even though they're called meat ants, they mostly eat honeydew from certain caterpillars and butterflies. In return, the ants protect the caterpillars from predators. They also scavenge for carcasses on the ground.

Meat ants are a dominant type of ant. Some ants try to avoid the meat ant. However, some species of ants fight with meat ants to establish territory.

Meat ants are able to kill poisonous cane toads. The cane toad's toxins are not likely to affect the meat ants. When a cane toad is attacked, it usually stays still and lets their poison kill their attacker. So, the ants are able to kill the toad while it just sits there.

In experiments, scientists used cat food to attract more meat ants to the ponds where cane toads breed. Almost no other species live in these areas. More than half of attacks by ants killed the toads immediately. Most of the toads that escaped died within 24 hours.

Meat ants don't often kill native toads. Many native toads go out at night instead of the day. Plus, native toads know to jump away from the meat ants.

Damage to other native organisms appears to be low, but scientists don't know for sure. Getting the ants to move to cane toads' ponds will likely lower the number of ants in other places. This could change the behaviors of other organisms.



The presence of meat ants across Australia has not prevented cane toads from invading that region, but meat ants do kill many toads. Increasing the numbers of ants where the toads breed could slow their spread.

Appendix M: Stage analysis coding manual

Coding Manual for Arguments in Second and Fourth Grade

Overall approach: Rate each applicable stage or stage pair as Successful (S), Partially Successful (P), or Unsuccessful (U) in terms of achieving the stage-specific goals presented in instruction. Broad descriptions of these goals are provided below, followed by grade level-specific descriptions and examples of writing at each of the three performance levels.

Description of Problem *“familiarizes the reader with important background on the issue or problem.”*

Claim is *“a statement that introduces the specific issue to be argued and the writer’s position on that issue.”*

Evidence is *“a piece of information from the text that supports the overall claim.”*

Reason is *“a statement about why/how the evidence supports the claim and interprets the evidence in relation to the larger issue.”*

Counterargument is *“a piece of information that someone presenting another side of the argument might use for evidence.”*

Evaluation *“evaluates the strength of the counterargument and explains what this information means for the writer’s position.”*

Restatement of Claim *“summarizes and restates the writer’s position on the issue.”*

2nd Grade

Stage/Stage Pair	Level	Description	Examples
<i>Evidence/Reason</i>	Successful	Presents a fact from the Grizzly Bears text and a logical explanation of how that information supports the chosen claim.	People are shooting the grizzly bears in the wild. They will be safer in zoos because people can’t kill them in zoos. 10 Grizzly bears need lots of space. If we put grizzly bears in zoos they won’t have space to move around. 45
	Partially Successful	A connection between evidence and claim is present but incomplete/insufficient, leaving more work to the reader than is ideal (“incomplete bridge”).	Grizzly bears need lots of space. In the zoo the cages are small. 18 Baby bears called cubs stay with their mother for 2 or 3 years. If we put bears in zoos the cubs will not know how to hunt. 63
	Unsuccessful	Lacks reference to Claim or anything that could otherwise be interpreted as a Reason, adds commentary that does not make sense, or includes Evidence and Reason that are not logically related.	Grizzly bears need a lot of space to live and hunt and play. Grizzly bears should be in zoos because they can get shot. 31 Grizzly bears are 7 feet tall. If we put grizzly bears in zoos because

			they will get poisonous fish. 124
<i>Counterargument/ Evaluation</i>	Successful	Provides a piece of information that someone presenting another side of the argument might use for evidence and explains why the Counterargument does not change the writer's position.	Some people say we should not put them in zoos because bears are 7 feet tall and zoo cages are small. I disagree because they can build a huge cages. 2 Some people say that bears should not be in zoos because they need a lot of space. I disagree because if they don't have some space they can build a new cage for the bears. 107
	Partially Successful	Presents Counterargument and attempts an Evaluation but does not adequately explain why the Counterargument does not change the writer's position (e.g., explanation is incomplete/logically weak).	Some people say that we can put grizzly bears in zoos, because they will be safe. I just disagree because grizzly bears don't like to be next to humans. 112 Some people say we should put grizzly bears in zoos because their habitat is shrinking. I disagree because the zoos are not the same size as 20 Oak Groves. 124
	Unsuccessful	Presents Counterargument but either does not include anything resembling an Evaluation or Evaluation is unrelated to Counterargument.	Some people say that they catch and eat salmon. But I say we could put bears in zoos. 128 Some people say that we should put grizzly bears in zoos because they are poisoning the water. I disagree because they need a lot of space. 72

4th Grade

Stage/Stage Pair	Level	Description	Examples
<i>Description of Problem/Claim</i>	Successful	Establishes that cane toads were a problem for Australia, explains why, introduces meat ants as a possible solution, and presents the writer's position.	The cane toads have become a huge problem in Australia. Their population has spread to about 2,000,000,000. They are highly poisonous so predators can't kill them. But scientists have discovered an organism that is not likely to be harmed by the toxins. The meat ant. This organism has many negative and positive things about it. Therefore we should study this more before we use it. 26 The cane toad invasion has

			become a big problem in Australia. They started off with 102 toads but had a big increase in population. 2,000,000,000 toads are now taking over northern Australia. They eat about anything that fits in their mouth and are endangering native predators. Scientists say using meat ants will help decrease the cane toad population by more than half the toads' population. Australia should not use it. 138
	Partially Successful	Issues vary, but generally one of the following applies: Provides a partial description of the issue but leaves out a key element; includes all the necessary elements but gives insufficient detail; adds details that were not strictly relevant and proved to be distracting; does not include Claim (but writer incorporates position elsewhere).	<p>The cane toads have been brought to solve a problem but they have become a bigger problem. They have been in Australia since the 1930s. Scientists think using meat ants is a good solution. In my opinion it is good but they should study. 58</p> <p>The cane toad has been a big problem to Australia. They were brought to the country to stop the beetles from eating sugar cane crops. But they didn't eat the beetles they ate anything that fits in their mouth. Scientists came up with an idea. They want to use meat ants. Meat ants are ants that eat meat. Meat ants are able to kill poisonous cane toads. The cane toads toxins are not likely to affect the meat ants. Also the cane toad mom can lay up to 35,000 eggs each year. 51</p>
	Unsuccessful	Omits a significant amount of information and leaves the reader unprepared to contextualize the subsequent elements of the arguments.	<p>Meat ants... I think it's a good option but should be studied more. 84</p> <p>The scientists in Australia think that maybe the meat ants are a really good idea but I think they should study more and if it works bad they should not use it no more. They found an organism to not be harmed by the toxins. 85</p>
<i>Evidence/Reason</i>	Successful	Provides a piece of information from the text that supports the overall claim and a statement about why/how the evidence supports the claim and interprets	One reason why scientists shouldn't use the meat ants is that even though they're called meat ants they mostly eat honeydew from caterpillars and

		<p>the evidence in relation to the larger issue.</p>	<p>butterflies. They also scavenge for carcasses. Why the meat ants aren't a good solution is that they're scavengers, which means they eat dead animals and are not used to hunting live prey. It may affect the ecosystem and organisms living there. 138</p> <p>One reason why meat ants are most likely a horrible solution is because meat ants mostly eat honeydew from caterpillars and butterflies. They also eat carcasses in the tremendous wild. The reason why scientists shouldn't use this information is because if the meat ants keep eating carcasses and honeydew it will likely affect their diet and maybe their health, also the organisms who eat these foods may not have anything to eat because meat ants are eating them and maybe the toad population will increase and Australia could have a very very huge problem. 116</p>
	<p>Partially Successful</p>	<p>A connection between evidence and claim is present but incomplete/ insufficient, or material following evidence does not function to make such a connection (e.g., reiterates claim or adds facts without interpreting).</p>	<p>One reason we should study this more before we use it is meat ants mostly eat honeydew from caterpillars. They also scavenge for carcasses on the ground. We should study this more because meat ants mostly eat carcasses and honeydew.</p> <p>Another reason we should study this more is cane toads rely on their toxins to kill their attacker. But the toxins are not likely to harm the meat ant. When an attacker attacks the toad, it just sits there and lets its poison harm the organisms. When the meat ant attacks, it can eat it easily because it is just sitting still. 26</p> <p>The first reason is that toxins from the toads are not likely to affect the ants. Also meat ants can kill the toads. The next reason is I think it's a good solution because the meat ants will probably keep on eating the toads and the toads problem will probably end. 94</p>

	Unsuccessful	Lacks reference to Claim or anything that could otherwise be interpreted as a Reason, or adds commentary that does not make sense.	One reason they should study the meat ants more is because they eat a lot of other things like carcasses and honeydew instead of cane toads. So, they should talk about it. A last reason is that meat ants only eat 2 cane toads in their lifetime. So only $\frac{1}{4}$ of the toads will die and $\frac{3}{4}$ will stay alive. 84
<i>Counterargument/ Evaluation</i>	Successful	Provides a piece of information that someone presenting another side of the argument might use for evidence and evaluates the strength of the counterargument, explaining why the Counterargument does not change the writer's position.	Some people may say that we should not use the meat ant because they could fight with other ants. Also they can change the behavior of other organisms. It is true that meat ants can start to fight with other ants for their territory because meat ants are the dominant type of ant. But they do not always fight. It is also true that they can change the behavior of other organisms. But it is not for sure that it is going to completely change the behavior. 26 If someone might argue with me that said it is bad, don't [use] it, they might say the meat ants eat honeydew from caterpillars and butterflies. So it will be hard to make them eat the cane toads. I agree that they eat honeydew, it is their main food. But I disagree that it will be hard to make them eat cane toads. They can put honeydew on the cane toads. 58
	Partially Successful	Presents Counterargument and attempts an Evaluation but does not provide sufficient context or does not adequately explain why the Counterargument does not change the writer's position (e.g., explanation is incomplete/logically weak, or concession to opposing position is excessive and does not support original position).	Some people might say meat ants can kill toads. The toxins aren't likely to hurt the ants. The toads just sit there. So the toad population could decrease. What if they don't eat the cane toads? We don't know if they will eat the toads, so the humongous toad population could get larger. 116 It is true meat ants can fight with many other ants. But they do not always fight other organisms. 85
	Unsuccessful	Presents Counterargument but either does not include anything resembling an Evaluation or adds commentary that does not further the argument (e.g., reiterates evidence or writer's	Some people say it's worth trying but in my opinion the scientists should study it more. 84 You may say that it is a good

		position on the issue, makes an unrelated statement).	idea but toxins from the toad aren't likely to affect the meat ant but the toad just sits there while the ant eats it. Also the toads die immediately. It may take 24 hours for the cane toads to die. This shows that meat ants are not a good solution because it might invade Australia after there's fewer toads. 138
<i>Restatement of Claim</i>	Successful	Summarizes and restates the writer's position on the issue.	Scientists are not sure if meat ants are a good solution. Therefore I think we should study it more before we use it. 26 Australia needs a good solution to this cane toad problem which is the meat ants but they should still study the meat ants and they should make sure meat ants are safe to the other animals and insects. 151
	Partially Successful	Restates the writer's position on the issue without additional summary.	This is why Australia shouldn't use the meat ants to help get rid of the cane toad. 138 That's what I think in my opinion. I think it's good but should be studied more. 84
	Unsuccessful	Includes a final statement but neither summarizes nor restates the claim.	On the other hand the cane toads have become a huge problem in Australia. By eating the big predators and small organisms. The cane toads hurt the meat ants. 85

Appendix N: Percentages of exact and adjacent agreement for primary trait scoring

2 nd Grade	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
MAIN						
Exact	48%	40%	32%	40%	44%	36%
Adjacent	32%	52%	60%	60%	40%	48%
Both	80%	92%	92%	100%	84%	84%
PRE						
Exact	20%	33%	32%	28%	40%	48%
Adjacent	44%	50%	28%	64%	60%	48%
Both	64%	83%	60%	92%	100%	96%
POST						
Exact	32%	32%	28%	32%	40%	40%
Adjacent	44%	52%	48%	64%	56%	52%
Both	68%	84%	76%	96%	96%	92%
Overall						
Exact	33%	35%	31%	33%	41%	41%
Adjacent	40%	49%	45%	63%	52%	49%
Both	73%	84%	76%	96%	93%	90%

N1. Inter-rater agreement for second grade texts

4 th Grade	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
MAIN						
Exact	37%	37%	41%	44%	26%	30%
Adjacent	48%	41%	30%	48%	52%	67%
Both	85%	78%	71%	92%	78%	97%
PRE						
Exact	44%	33%	23%	26%	33%	30%
Adjacent	52%	63%	48%	52%	56%	41%
Both	96%	99%	71%	78%	89%	71%
POST						
Exact	52%	37%	44%	59%	37%	26%
Adjacent	33%	44%	37%	30%	41%	56%
Both	85%	81%	81%	89%	78%	82%
Overall						
Exact	44%	36%	36%	43%	32%	38%
Adjacent	44%	49%	38%	43%	49%	54%
Both	88%	85%	74%	86%	81%	92%

N2. Inter-rater agreement for fourth grade texts

Appendix O: Independent Scores, Individual Results

Note: The first set of scores presented for each student are the final scores for the main argument. The second and third sets of scores are for the pre- and post-instruction writing, respectively. Final scores for each trait were derived by averaging scores provided by independent raters. “Sample” refers to the number randomly assigned to each piece of writing produced for this study.

Final scores for second-grade students:

Student Number	Sample	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions	
<i>Main 1</i>	75	5	4	5	4	4	4	
	<i>Pre</i>	83	4	3.67	4.33	4	4	2.67
	<i>Post</i>	39	5.67	5.33	5	4.33	5	4.33
2	107	5	4.5	5	4	4.5	4	
	24	4.5	4	4.5	4	3.5	3	
	118	3.33	3.33	3	3	2.33	2.33	
3	54	5	5	5	4.5	4	4	
	3	1.67	1.33	1.33	2	1.67	1.67	
	103	2	2.33	2	2.33	2	1.67	
4	153	3	2.33	2.67	2.67	2.33	2	
	12	2.5	2	2	1.5	1.5	1	
	134	3	2	2	2	1.5	1.5	
5	31	2.67	2.67	2.33	2.33	2.33	2.33	
	129	1.5	1.5	1.5	1.5	1.5	1.5	
	64	3	3	2	2.5	2.5	2	
6	63	4.67	3.67	4	3.67	3	2.67	
	146	3	3	3	3	2.33	1.67	
	130	4.33	3.67	4	3.33	3	1.67	
7	156	4.33	3.67	4.33	3	3.33	3	
	92	3.5	2.5	3	3	2.5	2.5	
	53	3.5	2.5	3.5	3.5	3	2	
8	2	4.67	4.33	4	4	4	4	
	80	2.33	2.33	2.33	3	3	3.67	
	27	2	1.5	1.5	3.5	2.5	3.5	
9	124	4	3.5	3.5	3.5	3	3	
	4	2.5	1.5	2	1.5	1.5	1	
	60	2.5	2.5	2	1.5	2	1.5	
10	128	4	3.33	3.33	3	2.67	2.33	
	70	2.67	1.67	2.33	2	1.22	1.67	

	7	1.5	1	1.5	1.5	1	1
12	10	5.5	4.5	4.5	4.5	4.5	4.5
	29	3.5	3	4.5	4	4	5
	147	5	4.5	4.5	4.5	3.5	4
13	91	4.67	5.67	5	4.33	4.33	5.33
	55	2.67	3.33	3	3	3	3
	32	4	3.5	4.5	4	4.5	4.5
14	86	3.67	4.33	3	3	3.33	3
	105	4	3	3	3	3	2.5
	102	3.33	3	3.67	3.67	3.67	2.67
16	78	5	5	5	4.5	5	4.5
	120	3	2.33	4	3.33	2.67	3.67
	33	4	3.67	4	3	4	3.67
17	45	4	4	3.5	3.5	3	3
	22	1.5	1.5	1.5	2.5	2	1.5
	66	2.67	2	2.33	3	2.33	1.67
18	71	5.5	4.5	5.5	4.5	4	3.5
	8	4	2.67	2.67	2.67	3	1.33
	5	4	3	3.5	3.5	3	2.5
19	106	2.33	1.67	1.67	2.33	2	1.67
	37	3	1.67	2.67	2.67	2.67	2
	44	2.5	2	2	2.5	2.5	1.5
20	74	4	2.5	3	3	3	2
	125	1.5	1.5	1	1.5	1.5	1
	144	2.67	2.33	2.67	2	2.33	1.67
21	72	4.33	4	3.33	3.67	3.33	3
	59	2.5	2	1.5	1.5	2	2
	149	3	2.5	2.5	2	2.5	1.5
22	98	4	3.33	3	3	3	1.33
	48	1	1	1	1.5	1	1
	21	3.33	2	2.67	2.33	2	1.67
23	18	5	4	5.5	4	4	3.5
	87	2.67	2.33	2	2.67	2.33	2.33
	14	3	2	2	2.5	2.5	2
24	88	4	3.5	3.5	3	3	2.5
	25	4	3	2.5	2	3.5	2.5
	97	2	2.33	1.67	2	2	1.33
25	112	4.5	4.5	5	4	4	3.5
	69	2.33	2	2.67	2.67	2.33	2.67
	141	3.5	2.5	2.5	3	2.5	2

26	140	5	4.5	4.5	4.5	3.5	3.5
	81	1	1	1	1.5	1	1.5
	89	1.5	1.5	1.5	1.5	1	1
27	123	4.67	4	4	3.33	3.33	3
	133	2.67	2.33	2.67	2.67	2.33	2.33
	38	2.5	2	2	1.5	1.5	1

Final scores for fourth-grade students:

Student Number	Sample	Ideas & Content	Organization	Voice	Word Choice	Sentence Fluency	Conventions
<i>Main</i> 1	94	3.5	3.5	3	3.5	3.5	3
<i>Pre</i>	73	1.5	2	1.5	2	1.5	2
<i>Post</i>	114	3	2.5	2	3	2.5	2.5
2	85	3.33	3	2.67	3.67	3.33	3.33
	35	2.67	2.67	2.67	2.67	2.67	3.33
	42	2.5	2.5	2	2.5	2	2
3	6	5	4	4	4	3.5	3.5
	152	4	4	3.5	3.5	3.5	3.5
	49	4	3.67	4	4	3.33	3.33
4	58	3.5	3	3	2.5	2	2
	52	1	1	1	1.5	1	1
	135	2.5	2	2	2	1.5	1
5	61	4	4	4.5	4	4	4
	43	3	2.5	2.5	3	1.5	1.5
	121	3	3	3.5	3	3.5	3.5
6	104	3	3.52	2.5	3	3.5	3.5
	9	3	1.67	2	2	1.67	2
	117	3.5	3.5	3	3	3	2.5
7	154	2.33	2.33	2.33	2.67	3	3
	99	2.5	2	2	2.5	2	2
	30	3	3.5	2.5	3	3	3
8	111	2.5	2	2	2.5	2.5	2
	110	3.67	3.33	3	3	3.33	3
	113	4.33	4	4.67	4	3.67	3.33
9	150	4	3.5	3	3.5	4	3.5
	142	3.5	3.5	3	3	3	3.5
	115	2.5	2.5	3	3	3	2.5
10	90	3.67	3	3.33	3.33	3	2.67
	62	2	2.33	1.33	2	2	2.33
	148	3	3	2.5	2.5	2.5	2.5

11	26	5.5	4.5	5	5.5	4.5	5
	132	4.67	4.67	5.67	4.33	4.33	4.67
	155	5	5	5.5	5	5	5
12	143	2.67	2	2	2.33	2	2
	41	2.5	2.5	2.5	3	3	3
	108	3.67	3.33	3.67	3.33	3.33	3.33
13	51	3	3	3	3.67	3	3.33
	11	4	3.5	3.5	3.5	4	4
	100	2.67	2.67	2	2.67	2.67	2
14	131	3.5	3.5	3.5	3.5	3.5	3.5
	126	5	4.67	5	4.33	4.67	4.67
	65	4.5	4.5	4.5	4	4.5	4.5
15	16	3	2.5	3.5	3.5	4	4.5
	93	4.33	4.33	4.67	4	4.33	4.33
	20	5	5	5	4.5	5	4.5
16	13	4.33	4.33	4	4	4	4
	34	4.67	4.67	4.33	4.67	4.33	5
	127	3	3	3	3	3.5	3.5
17	138	4.5	4	4	4	4.5	4.5
	136	3.5	3.5	3.5	3.5	3.5	3.5
	56	4.67	4	4.33	4	4.33	4.33
18	40	3	3	3	3.33	2.33	2.67
	82	2.5	2.5	2.5	2.5	2.5	2.5
	101	2.67	3	3	2.67	2.33	2.33
19	84	3.67	3.67	3.33	3.67	4	4
	79	3	3	3	3	3	2.67
	15	4	4	3.67	3.67	3.67	3.67
20	157	2	1.5	1	2	1.5	1.5
	145	2.5	2.5	2	2	2.5	2.5
	36	1.5	1	1	1.5	1	1
21	17	3.67	3.67	3.67	3.67	3.33	3.67
	96	2.5	2	2	2.5	2	2
	77	3	2.67	2.67	3	2.67	2.67
22	1	4.5	4.5	5	4	4.5	4.5
	46	4	3.33	3.67	4	3.67	3.33
	19	4	4	3.67	3.67	3.33	3.33
23	116	5.33	4.67	5.33	5	5	5
	119	3	3	3.5	3.5	3	3
	23	4.67	4.33	4.33	4.33	4.33	4.33

24	47	3	2.67	2.67	3.33	2.67	3
	109	2	2	2.5	2.5	2.5	2.5
	67	2	1.5	2	2.5	2	2
25	151	3.33	3.67	3	3.67	3.33	3.33
	57	3	3	2.67	3.33	3	3.33
	50	4	3.67	3.67	3.33	3.33	2.33
26	137	5	4	4	4	4	4
	139	3.67	3.67	3.33	3.33	3.67	3.67
	122	4	4	3.33	3.33	3.67	3.33
27	28	2	1.5	1.5	2	1.5	1.5
	95	2.5	2.5	2.5	2.5	2.5	2.5
	68	2	2	2	2	1.5	1.5

References

- Anderson, R.C., Chinn, C., Chang, J., Waggoner, M., & Yi, H. (1997). On the logical integrity of children's arguments. *Cognition and Instruction, 15*, 135-167.
- Anderson, R.C., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S.-Y., Reznitskaya, A., Tillmans, M. & Gilbert, L. (2001). The snowball phenomenon: Spread of ways of talking and thinking across groups of children. *Cognition and Instruction, 19*, 1-46.
- Andrews, R., Torgerson, C., Low, G., McGuinn, N., & Robinson, A. (2006). Teaching argumentative non-fiction writing to 7-14 year olds: A systematic review of the evidence of successful practice. Technical report. In *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. <http://eppi.ioe.ac.uk/reel>
- Andrews, R., Torgerson, C., Low, G., & McGuinn, N. (2009). Teaching argumentative non-fiction writing to 7-14 year olds: An international review of the evidence of successful practice. *Cambridge Journal of Education, 39*, 291-310.
- Anstrom, K., DiCerbo, P., Butler, F. A., Katz, A., Millet, J., & Rivera, C. (2010). *A review of the literature on Academic English: Implications for K-12 English Language Learners*. Arlington, VA: The George Washington University Center for Equity and Excellence in Education.
- August, D. & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Erlbaum.
- Bereiter, C. (1980). Development in writing. In L.W. Gregg & E.R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 73-93). Hillsdale, NJ: Lawrence Erlbaum.
- Block, M. K. (2013). *The impact of identifying a specific purpose and external audience for writing on second graders' writing quality*. (Order No. 3590410, Michigan State University). *ProQuest Dissertations and Theses* 139. Retrieved from <http://search.proquest.com.proxy.lib.umich.edu/docview/1432194852?accountid=14667>. (1432194852).
- Brisk, M., & Zisselsberger, M. (2010). We've let them in on a secret: Using SFL theory to improve the teaching of writing to bilingual learners. In T. Lucas (Ed.), *Preparing all teachers to teach English language learners* (pp. 111-126). Mahwah, NJ: Erlbaum/Taylor & Francis.
- Bruner, J. (1960). *The process of education*. Cambridge, MA: Harvard University Press.
- Cavagnetto, A.R. (2010). Argument to foster scientific literacy: A review of argument interventions in L-12 science contexts. *Review of Educational Research, 80*, 336-371.
- Chambliss, M.J. (1995). Text cues and strategies successful readers use to construct the gist of lengthy written arguments. *Reading Research Quarterly, 30*, 778-807.

- Chambliss, M.J., & Murphy, P.K. (2002). Fourth and fifth graders representing the argument structure in written texts. *Discourse Processes*, 34, 91-115.
- Chase, C.I. (1968). The impact of some obvious variables on essay test scores. *Journal of Educational Measurement*, 5, 315-18.
- Christie, F. (2012). *Language education: A functional perspective. Language Learning Monograph Series*. USA: Wiley-Blackwell.
- Christie, F., & Derewianka, B. (2008). *School discourse: Learning to write across the years of schooling*. London: Continuum.
- Christie, F., & Unsworth, L. (2005). Developing dimensions of an educational linguistics. In R. Hasan, C. M. I. M. Matthiessen & J. Webster (Eds.), *Continuing Discourse on Language: A functional perspective* (pp. 217-250). London: Equinox.
- Cohen, M., & Riel, M. (1989). The effect of distant audiences on students' writing. *American Educational Research Journal*, 26, 143-159.
- Coker Jr., D.L., & Erwin, E. (2011). Teaching academic argument in an urban middle school: A case study of two approaches. *Urban Education*, 46, 120-140.
- Coker, D., & Lewis, W.E. (2008). Beyond *Writing Next*: A discussion of writing research and instructional uncertainty. *Harvard Educational Review*, 78, 231-251.
- Common Core State Standards Initiative. (2011a). *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.
- Crowhurst, M. (1980). Syntactic complexity in narration and argument at three grade levels. *Canadian Journal of Education*, 5, 6-13.
- Crowhurst, M. (1987). Cohension in argument and narration at three grade levels. *Research in the Teaching of English*, 21, 185-201.
- Crowhurst, M. (1990a). Teaching and learning the writing of persuasive/argumentative discourse. *Canadian Journal of Education*, 15, 348-359.
- Crowhurst, M. (1990b). Reading/writing relationships: An intervention study. *Canadian Journal of Education*, 15, 155-172.
- Crowhurst, M., & Piché, G.L. (1979). Audience and mode of discourse effects on syntactic complexity in writing at two grade levels. *Research in the Teaching of English*, 13, 101-109.
- Culham, R. (2005). *6 + 1 traits of writing: The complete guide for the primary grades*. New York: Scholastic.
- Crammond, J.H. (1998). The uses and complexity of argument structures in expert and student persuasive writing. *Written Communication*, 15, 230-268.
- Derewianka, B. (1991). *Exploring how texts work*. Rozelle, NSW, Australia: Primary English Teaching Association.
- Duke, N.K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, 35, 202-224.
- Ellis, N., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics—Introduction to the special issue. *Applied Linguistics*, 27, 558-589.
- Englert, C.S., Raphael, T.M., 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.
- Farnan, N. & Dahl, K. (2003). Children's writing: Research and practice. In Flood, J.,

- Lapp, D., Squire, J.R., & Jensen, J.M. (Eds.), *Handbook of research on teaching the English language arts* (pp. 993-1007). Mahwah, N.J.: Erlbaum Associates.
- Felton, M., & Kuhn, J. (1997). The development of argumentative discourse skill. *Discourse Processes, 18*, 187-210.
- Ferretti, R.P., MacArthur, C.A., & Dowdy, N.S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology, 92*, 694-702.
- Flower, L.S. & Hayes, J.R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*, 365-387.
- Francis, D.J., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions*. Portsmouth, NH: Center on Instruction.
- Freedman, A., & Pringle, I. (1988). Why students can't write arguments. In N. Mercer (Ed.), *Language and literacy from an educational perspective, Vol. 2*. (pp. 233-242). Milton Keynes, UK: Open University Press.
- Fulkerson, R. (1996). *Teaching the argument in writing*. Urbana, IL: National Council of Teachers of English.
- García, E.E. and Cuéllar, D. (2006). Who are these linguistically and culturally diverse students? *Teachers College Record, 108*, 2220-2246.
- Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. London: Routledge.
- Gebhard, M., & Harman, R. (2011). Reconsidering genre theory in K-12 schools: A response to school reforms in the United States. *Journal of Second Language Writing, 20*, 45-55.
- Gebhard, M., Harman, R., & Seger, W. (2010). Reclaiming recess: Learning the language of persuasion. *Language Arts, 84*, 419-430.
- Genesee, F., Lindholm-Leary, K.J., Saunders, W., and Christian, D. (2006). *Educating English Language Learners*. NY: Cambridge University Press.
- Gibbons, P. (2002). *Scaffolding language, scaffolding learning: Teaching second language learners in the mainstream classroom*. Portsmouth, NH: Heinemann.
- Gilbert, J. & Graham, S. (2010). Teaching writing to elementary students in grades 4-6: A national survey. *The Elementary School Journal, 110*, 494-518.
- Graham, S., Bollinger, A., Booth Olson, C., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching elementary school students to be effective writers: A practice guide* (NCEE 2012- 4058). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch.
- Graham, S., & Harris, K.R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 323-344). New York: Guilford Press.
- Graham, S., Harris, K.R., & Mason, L.H. (2005). Improving the writing performance, knowledge, and motivation of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology, 30*, 207-241.
- Graham, S. & Perin, D. (2007). *Writing next: Effective strategies to improve writing of*

- adolescents in middle and high school.* Washington, DC: Alliance for Excellent Education.
- Graham, S., McKeown, D., Kiuahara, S., & Harris, K.R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*, 879–896.
- Graves, D.H. (1983). *Writing: Teachers and children at work.* Portsmouth, NH: Heinemann Educational.
- Halliday, M.A.K. (1994). *Introduction to functional grammar* (2nd ed). London: Edward Arnold.
- Halliday, M.A.K & Matthiessen, C. (1999). *Construing experience through meaning: A language-based approach to cognition.* New York: Cassell.
- Halliday, M.A.K. & Matthiessen, C. (2004). *An introduction to functional grammar* (3rd Ed.). New York: Arnold.
- Harris, K.R. , & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation.* Cambridge, MA: Brookline.
- Harris, K.R. , & Graham, S. (1999). Programmatic intervention research: Illustrations from the evolution of self-regulated strategy development. *Learning Disability Quarterly, 22*, 251-262.
- Harris, K.R., Graham, S., & Mason, L.H. (2006). Improving the writing, knowledge, and motivation of young struggling writers: The effects of self-regulated strategy development with and without peer support. *American Educational Research Journal, 43*, 295-340.
- Hillocks, G. (1999). *Ways of thinking, ways of teaching.* New York: Teachers College Press.
- Hillocks, G. (2002). *The testing trap: How state writing assessments control learning.* New York: Teachers College Press.
- Hillocks, G. (2010). Teaching argument for critical thinking and writing: An introduction. *English Journal, 99*(6), 24-32.
- Heath, S. B. (1983). *Ways with words.* New York, NY: Cambridge University Press.
- Hidi, S., Berndorff, D., & Ainley, M. (2002). Children’s argument writing, interest, and self-efficacy: An intervention study. *Learning and Instruction, 12*, 429-446.
- Horowitz, R. & Samuels, S.J. (Eds.) (1987). *Comprehending oral and written language.* San Diego: Academic Press
- Humes, A. (1983). Putting writing research into practice. *The Elementary School Journal, 84*, 3-17.
- Hunt, K. (1965) *Grammatical structures written at three grade levels. Research Report 3.* Urbana, IL: NCTE.
- Jadallah, M., Anderson, R.C., Nguyen-Jahiel, K., Miller, B.W., Kim, I.-H., Kuo, L.-J., Dong, T., & Wu, X. (2011). Influence of a teacher’s scaffolding moves during child-led small-group discussions. *American Educational Research Journal, 48*, 194-230.
- Jeong, J. S., Gaffney, J. S., & Choi, J. O. (2010). Availability and use of informational text in second, third, and fourth grades. *Research in the Teaching of English, 44*, 435-456.
- Kamberelis, G. (1999). Genre development and learning: Children writing stories, science reports, and poems. *Research in the Teaching of English, 33*, 403-460.
- Kamberelis, G. & Bovino, T.D. (1999). Cultural artifacts as scaffolds for genre

- development. *Reading Research Quarterly*, 34, 138-170.
- Knudson, R.E. (1992a). Analysis of argumentative writing at two grade levels. *Journal of Educational Research*, 83, 169-179.
- Knudson, R.E. (1992b). The development of written argumentation: An analysis and comparison of argumentative writing at four grade levels. *Child Study Journal*, 22, 167-184.
- Kuhn, D. (2005). *Education for thinking*. Cambridge: Harvard University Press.
- Lindholm-Leary, K., & Borsato, G. (2006). Academic achievement. In F. Genesee, K. Lindholm-Leary, W.M. Saunders and D. Christian (Eds.), *Educating English language learners: A synthesis of research evidence* (pp. 176-222). New York, Cambridge University Press.
- Loban, W. (1976). *Language development: Kindergarten through grade twelve. Research Report 18*. Urbana, IL: National Council of Teachers of English.
- Luke, A. (1996). Genres of power? Literacy education and the production of capital. In R. Hasan & G. Williams (Eds.), *Literacy and society* (pp. 308-338). London: Longman.
- Lunsford, A.A., & Ruskiewicz, J.J. (2004). *Everything's an argument, 3rd edition*. Boston: Bedford/St. Martin's Press.
- Macken-Horarik, M. (1998). Exploring the requirements of critical school literacy: A view from two classrooms. In F. Christie & R. Mission (Eds.), *Literacy and Schooling* (74-103). New York: Routledge.
- Malvern, D., Richards, B., Chipere, N., and Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke, U.K.: Palgrave Macmillan.
- Markham, L.R. (1976). Influence of handwriting quality on teacher evaluation of written work. *American Educational Research Journal*, 13, 277-83.
- Martin, J. R. (1989). *Factual writing*. Oxford: Oxford University Press.
- Martin, J. R. (2009). Genre and language learning: A social semiotic perspective. *Linguistics and Education*, 20, 10-21.
- Martin, J. R., & Rose, D. (2003). *Working with discourse: Meaning beyond the clause*. London: Continuum.
- Martin, J. R., & Rose, D. (2008). *Genre relations: Mapping culture*. Oakville, CT: Equinox.
- Martin, J.R. & White, P.R.R (2005). *The language of evaluation: Appraisal in English*. New York: Palgrave Macmillan.
- McCann, T.M. (1989). Student argumentative writing: Knowledge and ability at three grade levels. *Research in the Teaching of English*, 23, 62-76.
- McCarthy, S.J. (2008). The impact of No Child Left Behind on teachers' writing instruction. *Written Communication*, 25, 462-505.
- Means, M.L., & Voss, J.F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction*, 14(2), 139-178.
- Midgette, E., Haria, P., & MacArthur, C. (2008). The effects of content and audience awareness goals for revision on the persuasive essays of fifth- and eighth-grade students. *Reading and Writing*, 21, 131-151.

- Michaels, S., O'Connor, C., & Resnick, L. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education*, 27, 283-297.
- Moore, J., & Schleppegrell, M. (accepted). Using a functional linguistics metalanguage to support academic language development. *Linguistics and Education*.
- Myhill, D. (2007). Becoming a designer: Trajectories of linguistic development. In R. Beard, D. Myhill, J. Riley, & D. Nystrand (Eds.), *Handbook of writing development* (pp. 402-414). London: Sage.
- Myhill, D. (2008). Towards a linguistic model of sentence development in writing. *Language and Education*, 22, 271-288.
- NAEP. (1994). *The National Assessment of Educational Progress 1992 Report Card*. Princeton, NJ: Educational Testing Service.
- NAEP. (1999). *National Assessment of Educational Progress Writing Report Card for the Nation and the States*. Retrieved from <http://nces.ed.gov/nationsreportcard/pdf/main1998/1999462.pdf>
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards*. Washington, DC: Authors. *Appendix A: Research supporting key elements of the standards*. Retrieved from http://www.corestandards.org/assests/Appendix_A.pdf
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66, 60-92.
- Newell, G.E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly*, 46, 273-304.
- O'Hallaron, C. (2012, April.) *Supporting elementary-level ELLs' writing development with FG metalanguage*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, BC.
- Pappas, C.C. (1993). Is narrative "primary"? Some insights from kindergartners' pretend readings of stories and information books. *Journal of Reading Behavior*, 24, 97-129.
- Perera, K. (1984). *Children's writing and reading: Analyzing classroom language*. New York: Blackwell.
- Perie, M., Grigg, W., Donahue, P. (2005). *The nation's report card: Reading 2005* (NCES 2006-451). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Purcell-Gates, V., Duke, N.K., & Martineau, J.A. (2007). Learning to read and write genre-specific texts: The roles of authentic experience and explicit teaching. *Reading Research Quarterly*, 42, 8-45.
- Read, S. & Landon-Hays, M.M. (2013). The knowing/doing gap: Challenges of effective writing instruction in high school. *Teaching/Writing: The Journal of Writing Teacher Education*, 2(2), 6-15.
- Reznitskaya, A., Anderson, R.C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discourse on written argument. *Discourse Processes*, 32, 155-175.
- Reznitskaya, A., Anderson, R.C., & Kuo, L-J. (2007). Teaching and learning argumentation. *The Elementary School Journal*, 107, 449-472.
- Reznitskaya, A., Kuo, L-J., Glina, M., & Anderson, R.C. (2009). Measuring

- argumentative reasoning: What's behind the numbers? *Learning and Individual Differences*, 19, 219-224.
- Rex, L.A., Thomas, E.E., & Engel, S. (2010). Applying Toulmin: Teaching logical reasoning and argumentative writing. *English Journal*, 99(6), 56-62.
- Rothery, J. (1996). Making changes: Developing an educational linguistics. In R. Hasan & G. Williams (Eds.), *Literacy in society* (pp. 86-123). New York: Longman.
- Scardamalia, M., & Bereiter, C. (1983). The development of evaluative, diagnostic, and remedial capabilities in children's composing. In M. Martlew (Ed.), *The psychology of written language: Developmental and educational perspectives* (pp. 67-95). London: John Wiley & Sons.
- Schleppegrell, M.J. (2004). *The language of schooling*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Schleppegrell, M.J. (2008). Grammar, the sentence, and traditions of linguistic analysis. In Charles Bazerman (Ed.) *Handbook of Writing Research* (pp. 549-564). Mahwah, NJ: Erlbaum.
- Schleppegrell, M.J., & de Oliveira, L. (2006). An integrated language and content approach for history teachers. *Journal of English for Academic Purposes*, 5, 254-268.
- Schleppegrell, M.J., Moore, J.P., Al-Adeimi, S., O'Hallaron, C.L., Palincsar, A.P., & Symons, C. (in press). Tackling a genre: situating SFL genre pedagogy in a new context. In Luciana de Oliveira and Joshua Iddings (Eds.), *Genre Pedagogy across the Curriculum: Theory and Application in US Classrooms and Contexts*. Sheffield, UK: Equinox.
- Schleppegrell, M. J. & O'Hallaron, C.L. (2011). Teaching academic language in L2 secondary settings. *Annual Review of Applied Linguistics*, 31, 3-18.
- Schulze, J. (2011). Writing to persuade: A Systemic Functional view. *Gist Education and Learning Research Journal*, 5, 127-157.
- Snow, C.E. & Kim, Y.-S. (2010). Syntactic and discourse structures: What should fourth and fifth graders know and be able to do? Briefing paper prepared for the New Standards Committee on 4th and 5th grade literacy, National Center on Education and the Economy and the University of Pittsburgh.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Swales, J., & Feak, C. (2004). *Academic writing for graduate students: Essential tasks and skills (2nd ed.)*. Ann Arbor: University of Michigan Press.
- Toulmin, S.E. (1958). *The uses of argument*. New York: Cambridge University Press.
- Varghese, S.A. & Abraham, S.A. (1998). Undergraduates arguing a case. *Journal of Second Language Writing*, 7, 287-306.
- Ventola, E. (1997). Modalization: Probability—an exploration into its role in academic writing. In A. Duszak (Ed.) *Culture and styles of academic discourse* (pp. 157-179). New York: Mouton de Gruyter.
- Voss, J.F., & Means, M.L. (1991). Learning to reason via instruction in argumentation. *Learning and Instruction*, 1, 337-350.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.) Cambridge, MA: Harvard University Press.
- Warren, J.E. (2010). Taming the warrant in Toulmin's model of argument. *English*

- Journal*, 99(6), 41-46.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency accuracy and complexity*. Honolulu, HI: University of Hawai'i Press.
- Wong Fillmore, L., & Snow, C.E. (2000). *What teachers need to know about language*. Washington, DC: Department of Education.
- Yeh, S. (1998). Empowering education: Teaching argumentative writing to cultural minority middle-school students. *Research in the Teaching of English*, 33, 49-83.