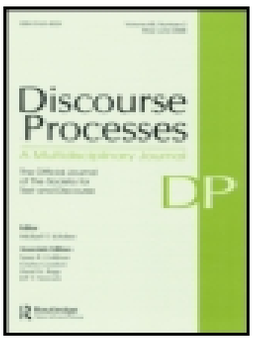

Discourse Processes

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hdsp20

Arguing for Teachers and for Friends: Eighth-graders' Sensitivity to Argumentation Features When Judging and Revising Persuasive Essays

Lisa B. Hsin & Catherine E. Snow

To cite this article: Lisa B. Hsin & Catherine E. Snow (2020): Arguing for Teachers and for Friends: Eighth-graders' Sensitivity to Argumentation Features When Judging and Revising Persuasive Essays, Discourse Processes, DOI: 10.1080/0163853X.2020.1803032

To link to this article: https://doi.org/10.1080/0163853X.2020.1803032

Published online: 01 Nov 2020.

DISCOURSE PROCESSES
https://doi.org/10.1080/0163853X.2020.1803032

Routledge
Taylor & Francis Group

Arguing for Teachers and for Friends: Eighth-graders' Sensitivity to Argumentation Features When Judging and Revising Persuasive Essays

Lisa B. Hsin and Catherine E. Snow

Harvard University

ABSTRACT

A novel instrument, the Features of Excellent Arguments task (FEXA), was developed to elicit adolescents' judgments about argumentative essays displaying to varying degrees features characteristic of strong persuasive writing: academic language, rich evidence, multiple perspectives, and rhetorical appeal. We collected students' categorical choices about the purposes to which the essays were suited (e.g., convincing a friend, impressing a language-arts teacher), discussions of those choices, and a revision of their selected "least persuasive" essay. In this first study using the FEXA, we found that the essays judged most persuasive largely overlapped with those judged most likely to earn a good grade (i.e., those with rich evidence and academic language). Students explained their preferences by mentioning formal language, use of evidence, and essay structure, while their revisions of their "least persuasive" selections focused on linguistic features and on adding or elaborating reasons. Implications for argumentative writing instruction and assessment are discussed.

Introduction

Skill in argumentation is widely considered central to Western civil discourse, even if its persuasive power is often unrealized (e.g., Kuhn et al., 2020). Yet only in the last decade or so has American education placed a heavy emphasis on teaching it orally or in writing (e.g., National Governors Association Center for Best Practices & Council of Chief State School Officers [NGAC], 2010). The argumentation literature has described age/grade differences in argumentation skill (e.g., Felton, 2004; Kuhn et al., 2016) as well as the influences of modality, topic, and data sources on how a person produces or evaluates arguments (see Voss & Dyke, 2002). But what is not yet well understood is how students themselves attend to and value various approaches to argumentation (but see Kuhn et al., 2010). As the pressure grows to write well, persuasively, and for diverse audiences, it is important to determine what students *believe* about the features and functions of written arguments and what consequences those beliefs might have for their writing activities.

The current study explores adolescents' judgments of and beliefs about written argumentative essays, as well as their skill in revising them, using the new "Features of Excellent Arguments" task (FEXA). Rather than asking students to write an argument "cold," with no preparation, or "warm," with some evidence provided but little engagement, the FEXA was designed to be a "hot" task (see Hsin, Phillips Galloway, & Snow, under review). That is, it recruits the enjoyment adolescent students derive from rendering "expert" judgments and criticisms of existing argumentative texts on familiar topics, especially when those judgments involve consideration of purpose and audience (e.g., handing in for a grade vs. entertaining a friend). The emphasis on arguments' suitability for academic and

CONTACT Lisa B. Hsin l.b.hsin@gmail.com; lisa_hsin@gse.harvard.edu Harvard University

social purposes is a key innovation of the FEXA: The social context of argument is sometimes included as an experimental manipulation (e.g., Mayweg-Paus et al., 2016), but it is rarely the subject of explicit participant reflection. To begin, we briefly review key literature on what characterizes good written arguments, on students' sensitivity to writing for different purposes, and on the uses of peer assessment and revision in learning how to write arguments.

Features of good persuasive argumentation

Although grammar, logic, and rhetoric—skills essential to argumentation—formed the trivium of classical education (McLuhan, 2009), the more recent history of writing instruction in U.S. schools has been largely dominated by narrative genres (Cutler & Graham, 2008), with only occasional excursions into argumentative essays. In those excursions, the terms "persuasive argument(ation)," "persuasive writing," and "argumentation" have been variously used. In this article we reproduce the terms that researchers themselves used when reviewing other literature, and ourselves refer to "written persuasive arguments," following Crowhurst (1990): "writing that takes a point of view and supports it with either emotional or logical appeals [or both]" (p. 349).

The standard characterization of written persuasive arguments is centrally composed of the point of view adopted in an essay (stance/position) and the nature of the support (evidence) for that point of view, as well as how they are linked (warrant; e.g., Crowhurst, 1990; Kuhn & Udell, 2007; Toulmin, 1958). This characterization ignores that much effective persuasion is not an argument as such, in the classical logical sense. That is, people are often persuaded by feelings of co-membership, by stances publicly taken, by the suggestion of expertise, by a sense of scarcity, and other factors that go beyond or beside argument (e.g., Cialdini, 2004; Cialdini & Goldstein, 2004). However influential these factors may be in "real-world" persuasion, the school-based standards for persuasive argumentative writing compel a more "cognitive," conceptual approach to the construction of written arguments. What, then, are the key features of argumentation that secondary instruction should support? We argue for four central domains: basic argument structure, perspective taking, academic language, and rhetorical appeal.

Basic argument skills include warranting claims and using evidence (Kuhn & Franklin, 2006). Acquiring persuasive argumentative skill requires that an arguer bring to bear facts about the world that can support her position, whether these are facts received on authority (e.g., Gelfert, 2011) or, more reliably, recognized by a community of experts (e.g., Maloney & Simon, 2006). Adolescents as young as middle-school age are capable of using, and even assessing the role of, evidence in support of arguments (Kuhn et al., 2013). Their argument metatalk confirms that they are aware of the importance of supplying evidence for claims (Kuhn et al., 2013). Most research on argumentative skill development focuses on evidence use (and often its relation to pro and con claims; e.g., Zohar & Nemet, 2001). Indeed, an argument without rich evidence is not likely to garner approbation from teachers or from scorers assessing essays produced for state accountability purposes.

Social dimensions of argumentative competence (see Kuhn et al., 2013) include the consideration of multiple perspectives. Often an arguer only considers "her side" of an argument, that is, her own perspective (Kuhn & Crowell, 2011). Adolescents are less adept than adults at coordinating attention to more than one position in an argument, which appears to result from their inattention to other positions rather than from incomprehension of them (Kuhn & Udell, 2007). But bringing multiple points of view to bear on an argument can help students meet two key requirements of successful argumentative discourse (Felton & Kuhn, 2001): understanding discourse goals, and applying effective strategies to meet them.

A third feature that is key to good persuasive argumentation, but often overlooked, is the suite of requisite academic language skills—skills in using the linguistic features of formal, academic discourse. The academic language register is customarily characterized by a detached or authoritative stance, conciseness and information density, explicit discourse organization, abstract lexicon, and complex syntax (Snow & Uccelli, 2009; Uccelli et al., 2015). As such,

academic language can support the explicit structure of an argument; clarify relationships among claims, counterclaims, evidence, and other argument components; facilitate the textual conciseness through which ample evidence can be conveyed; and project an authoritative voice (Snow & Uccelli, 2009). Discourse connectives (e.g., Basturkmen & von Randow, 2014), sophisticated word choice (e.g., Crossley et al., 2014; Olinghouse & Wilson, 2012), and more complex syntax (e.g., Crossley et al., 2011) have been confirmed as aspects of sophisticated argumentation. But more study is needed to understand the contributions of the other academic language features mentioned here to written argumentation quality.

Finally, arguments are likeliest to persuade when they are written with their audience in mind, that is, when they exhibit rhetorical appeal. Readers tend to respond to written persuasive arguments based in large part on their rhetorical rather than their "logical" features (Deane, 2013; Perelman, 2012). The rhetorical stance characteristic of academic writing is detached and authoritative, but this may elude or deter even those students who grasp the basic structure of sound arguments. A persuasive argument that embodies a more hortatory rhetorical style may hold considerable appeal to adolescents, especially given the frequency with which it is modeled in public discourse (see Melzer, 2009). That is, good arguments and successful arguments are not necessarily the same, and rhetorical appeals might lead an argument to succeed even though it "fails" on formal criteria.

In ideal persuasive argumentative writing, rich evidence, multiple perspectives, academic language, and rhetorical appeal would all conspire to produce a maximally effective argument. But still-developing adolescents may not be able to use all these features simultaneously (e.g., Anderson et al., 1997; Clark & Delia, 1976; Felton, 2004; Felton & Kuhn, 2001; Scardamalia & Bereiter, 1987); furthermore, they may not believe they are equivalently useful across all contexts of persuasive argument production or comprehension. Therefore, in the current study we sought to understand what students believe makes a good persuasive argument by asking them to read and evaluate essays differentially characterized by each of these features. Students rendered forced-choice judgments about the appropriateness of those varied essays for specific purposes, in particular a more academic one and a more social one.

Different goals for writing arguments

By attending to and manipulating the range of features that influence an argument's persuasiveness, readers and writers can take on various goals when reading or producing written arguments (Nussbaum, 2005). Providing subgoals based on elements of argumentative discourse has been found to improve fourth- and sixth-grade students' written argumentation (Ferretti et al., 2009) as well as their revisions of their own written arguments (Ferretti et al., 2000). The resulting essays in those studies showed an increase in overall persuasiveness and a greater frequency of specific sophisticated argument moves like the rebuttal of an alternative position. Similarly, when fifth- and eighth-grade students were assigned the goal of persuading an audience when revising their argumentative texts, that goal influenced the revisions they undertook, leading to improvements such as more consideration of opposing positions (Midgette et al., 2008).

An emphasis on the purpose of writing arguments has made its way into the teaching-focused literature, where the structures and functions of argumentation are contrasted with the characteristics of writing in other genres (e.g., Lapp & Fisher, 2012). And with only a few exceptions (e.g., Nussbaum, 2005), studies that have explored how to increase the persuasiveness of students' argumentative writing have done so presupposing a unitary purpose for the teaching of writing in this genre. Written argumentation, however, plays many roles in our society, and the form an argument should take varies by context (Deane, 2013). While we know that students read texts differently according to the task and context set for the reading activity (Rouet et al., 2017), it is unclear whether they can view a single already-written text through different lenses, flexibly, as they consider its appropriateness for different purposes. The literature on peer assessment of writing offers some indication of how students might approach such tasks.

Skill in peer assessment of writing

Students' tendency to write and read differently depending on their goals is also manifested when they are asked to *assess* writing, especially writing produced by their peers. Peer assessment can vary along numerous dimensions, including the range of instructions given to the student in the role of assessor, how many students are assessing a single written product and how many written products a given student assesses (e.g., Cho et al., 2006), the social relationships among assessors and assessed (e.g., Strachan & Wilcox, 1996), the personalities of those involved (e.g., Carrell, 1995), whether the work is also assessed by an expert or instructor (e.g., Kaufman & Schunn, 2011), and so forth (see Topping, 2010). Overall, meta-analyses have shown peer assessment and feedback to be effective for helping the original writers improve their assessed written products (Andrews et al., 2009; Graham & Perin, 2007; Hoogeveen & van Gelderen, 2013).

While most of this research focuses on how peer assessment affects the work of the original writer (e.g., Tsagari & Meletiadou, 2015), some studies have also explored how engaging in peer assessment affects the student doing the assessing. Such analyses often focus on revision that is undertaken in pairs, in which case the original writer also provides feedback on her classmate's work (e.g., Duran Gisbert & Monereo Font, 2008). The feedback that peers provide can vary dramatically in content as well as structure, and a variety of factors influence whether others' texts are rated positively or negatively (Strijbos & Sluijsmans, 2010). Like the first-language writers studied in most of this research, second-language learners are also capable of giving peers useful feedback on writing in both narrative and persuasive genres, and they are likewise able to use that feedback to revise their own writing (Villamil & Guerrero, 1998). In the current study we use students' proclivity for assessing peers' writing to elicit their judgments about persuasiveness in written arguments. Yet we also avoid the negative affect sometimes associated with judging close peers' writing or with receiving judgments rendered by peers, by presenting those essays as if written by unfamiliar peers from a neighboring school.

Improving writing with revision

The most common form of peer assessment may be the assignment of a score, often with some written feedback, but perhaps the most common form of *self*-assessment is the practice of revision. Much research on students' textual revisions has been focused on college-level students (e.g., Faigley & Witte, 1981; Fitzgerald, 1987; Zimmerman & Kitsantas, 2002) and in many cases on English-as-a-Foreign/Second-Language instruction (e.g., Mendonça et al., 1994; Min, 2006; Spycher, 2007; Villamil & Guerrero, 1998). But revision, and critical evaluation of writing more generally, represents an understudied component of metacognitive activity that is relevant to and attainable by much younger students as well (Fitzgerald & Markham, 1987; Graham & Perin, 2007). As we have already seen, some work has manipulated the revision process to help adolescent students structure their writing for certain goals (e.g., Ferretti et al., 2000; Midgette et al., 2008). Furthermore, exposure to instruction in oral argumentation improves the revisions that adolescents make to their own argumentative writing, leading to substantive revisions that make their arguments more persuasive (Early & Saidy, 2014). That is, students can internalize norms of argumentation in the spoken modality—even though it is dialogic, rather than monologic—and apply them to improving their own argumentative writing.

It remains to be seen whether those internalized norms can also be applied when students are asked to revise writing that is not theirs, and when the revisions have no implications for the students' own writing success. On one hand, we might predict that students will not revise as thoroughly when working with someone else's writing because the motivation to do so could be more limited, and because the need to take the perspective of the work's original author requires both social sensitivity and metacognitive resources. On the other hand, however, a critical stance may be easier to adopt when judging someone else's writing (cf. Cho et al., 2006), and the cognitive load may be reduced by the presence of a text defined as being subject to improvements. Revising others' writing may thus give

a more accurate measure of students' literacy skills, including their perspective-taking skills, than the revision of their own work where defensiveness and myside bias might inhibit criticism (e.g., Cho & MacArthur, 2010; Strijbos et al., 2010; Villamil & Guerrero, 1998).

Current study

In light of widespread adoption of Common-Core-inspired curricula (NGAC, 2010) and the long-term goal of ensuring all students have access to mature, useful argumentative practices (Graff, 2003), it is imperative that the research and practitioner communities be informed about students' understandings of what good written argumentation looks like and what purposes it serves. The field has established informed pedagogical targets, and innovative ways of assessing students' relationship to those targets are making their way into researchers' and teachers' hands (Deane et al., 2018). But we lack information about what students themselves perceive to be good persuasive arguments—and whether their perceptions of persuasiveness align with their perceptions of the instructions and examples they receive in school, which in turn should shape curriculum and instruction.

We undertook to design and administer an assessment that would provide data and insights about students' beliefs about the effectiveness of written arguments and their skills in judging and revising arguments. This report addresses the following research questions (RQs):

1. What do students' judgments of strategically varied written arguments reveal about the value they attach to evidence, academic language, counterarguments, and rhetorical appeals? Are their preferences related to their backgrounds or scores in ELA?
2. How do students explain the choices they make about which essay would get the best grade and which essay would be most fun to read?
3. What features do students introduce into, or remove from, essays that they judge to be less persuasive when asked to improve them, and how to these relate to their explicit judgments?

In this initial study using the FEXA approach, we hypothesized that students' opinions would reflect features commonly emphasized in the classroom—such as specific features of the prototypical five-paragraph essay—as well as features more characteristic of strong arguments in the real world—such as providing plausibly reliable evidence or use of rhetorical appeals.

Methods

We used an iterative process to design a novel instrument, the FEXA. The FEXA elicits students' judgments of argumentative essays emphasizing different features characteristic of strong persuasive writing: academic language, rich evidence, multiple perspectives, and rhetorical appeal. In asking students to make such judgments, we collected not only categorical choices but also explicit discussions of students' reasons for their choices and finally the revision of the essay that students found weakest.

Setting and participants

The FEXA was developed with and ultimately administered among students and teachers implementing Word Generation Weekly (WG: Snow et al., 2009; wordgen.serpmedia.org), a Tier-1 cross-disciplinary discussion-and-debate based curriculum. WG regularly exposes students to dialogic/group discussion and opportunities for written argumentation, but almost all argumentation-focused guidance is directed toward dialogic or whole-class debates, not extended written arguments (which are themselves not a focus of the curriculum). There is only sparse guidance in WG for the features of argumentation studied here. Nevertheless, these features are latent throughout the curriculum, as WG was designed to teach academic language and deep reading comprehension through

Table 1. Sample Demographic and Performance Characteristics

Group		*n*	%
Female		83	51
Limited English proficient		4	2
Has individualized education program (IEP)		30	19
PARCC	Met expectations	31	19
	Approached expectations	95	59
	Did not or partially met expectations	27	17
	Total	162	

Performance ratings are reflective of state-provided PARCC-ELA scores (see http://www.parcconline.org/; number of performance levels does not sum to the size of the whole sample due to missing data).

engagement with diverse sources of evidence and multiple perspectives on issues. Beyond these aspects of the curriculum we cannot be certain of classroom practices or what emphases teachers may have placed on the features of academic or argumentative writing. However, conversations with students and artifacts visible in the classrooms suggested that instruction in written argumentation focused on surface-level structural features (e.g., ordering of claims and reasons, avoidance of the first person). The participants were 162 eighth-grade students in a semi-urban middle school in the northeastern United States that was actively implementing WG. The demographic and performance characteristics of the sample are presented in Table 1.

Materials

Students participating in the FEXA read four essays arguing for the same side of a main claim in reaction to a contentious question and then answered several questions eliciting their opinions about those essays. Specifically, students responded to questions about their prediction of which essay would earn the highest grade in ELA class (Grade) and which essay would be most interesting for a close friend to read (Friend). These two questions were designed to elicit differing responses from students to whom the contrast between formal argumentation (the features emphasized in their instruction) and rhetorical and interpersonal appeals (those likely to appeal to peers) was accessible.

The participants were also asked which essays were most persuasive (+Persuasive) and least persuasive (–Persuasive). Participants were given a working definition of "persuasiveness" in the relevant prompt: "Which essay do you think is the most persuasive? That is, if you didn't already agree with the author's position, which one would be likeliest to bring you over to her/his side? Or, if you did already agree, which one did you think made the best case for the position?" While there are many ways of defining persuasiveness, this characterization was in line with what teachers reported conveying to their students and reflects one of the possibilities found in the literature.

Students were also asked to explain why their choices of essays for the Grade and Friend questions were the same or different in an "Explanations" question. Finally, the FEXA invited students to revise and improve the essay they had found least persuasive: "Show the author how you can make the same points that he or she made, but in a better way." In this component of the task, we did not reintroduce the term *persuasive* because we did not want to encourage students to introduce new argument components (claims/reasons) but rather expected to elicit key structural and rhetorical modifications. The materials emphasized that no one response was correct ("There are no right or wrong answers—we want to know what you think!").

Two forms were created, one addressing the topic of recycling ("Should recycling be required by the city?") and the other the topic of athletic uniforms ("Should student athletes design their own uniforms?"). The two topics were chosen to be similar to, but not overlap with, WG unit topics and to be of broad likely interest among middle-school students. We reasoned that these topics reflected slightly different normative queries, given that recycling requires virtually no particular individual talent and can be implemented by everyone, while uniform-design could rely on artistic talent, which

varies across individuals. We expected that this normative difference might lead to different preferred persuasive strategies among participating students.

For each of the two topics, four mock student essays were created based on real samples of middle-schoolers' responses to similar questions (using data from students in the treatment condition of a randomized controlled trial of WG Weekly, none of whom participated in the current study) and were labeled with gender-neutral "author" names. Each essay type was characterized by emphasis on one set of features: academic language, rich evidence, multiple perspectives, or rhetorical appeal. The feature sets were developed with reference to the literature and in consultation with master's students in language and literacy as well as middle-school teachers. Example sentences reflecting each of the four feature sets can be found in (1).

(1) a. *Academic language*: "The city's policies could stay in place for many years, and as a result, the environment would be left in better condition to sustain future animal and plant life."
b. *Rich evidence*: "For one, scientific data shows that the planet is warming and extra trash is at least partly to blame."
c. *Multiple perspectives*: "Yet city leaders could also encourage the citizens to be committed to the city's condition in other ways, because they too have experienced both being required and feeling personally driven to take care of it."
d. *Rhetorical appeal*: "Last year the city just permitted throwing anything we didn't need into the garbage, but that was pointlessly wasteful, so they made a policy that said that we would have to recycle, and we would get an award if we did it a lot."

A norming study confirmed that adult readers with experience in teaching adolescents ($N = 14$) overwhelmingly matched the essay types with their target features (academic language, 75%; rich evidence, 79%; multiple perspectives, 82%; rhetorical appeal, 89%). Those who did not respond as expected tended to exchange the academic language and rich evidence categories with one another and occasionally the academic language category with multiple perspectives.

Across the forms and feature sets, each essay was designed to be approximately equivalent in length (in Recycling essays, mean length = 114.25 words, SD = 9.50; in Uniforms essays, mean length = 117.25 words, SD = 6.29; $t(3) = 0.53$, $p = .62$). Each pair of essays across forms reflecting the same feature (e.g., the "academic language" essay on recycling and the "academic language" essay on uniforms) was created from an essay template such that the sentential and discourse structure of the paragraphs was virtually identical. We confirmed the linguistic parity of the essays by computing their Coh-Metrix L2 Readability scores, which likewise did not differ significantly between the forms within each of the essay feature pairs (paired $t(3) = 1.73$, $p = .18$).

Evidence from FEXA pilot

The FEXA was originally developed and piloted in consultation with a separate group of teachers and students, who were also simultaneously implementing WG. The pilot version differed from the final version in several important respects, but the results of the pilot data were largely consistent with the data reported here and therefore offered a first suggestion of the validity of the instrument. Aside from some wording improvements, the FEXA pilot differed from its final form in that (1) it was administered on paper rather than electronically, (2) it distributed three distinct sets of essays among participants rather than two, (3) the three sets of essays did not systematically share sentence-/discourse-structure features, and (4) the revision task was implemented across students in one of two ways, either by soliciting in-line revisions or a fresh rewrite.

The pilot was conducted with a sample of 140 students in a middle school in a semi-urban district the northeastern United States. Teachers in the school volunteered to administer the task to one or more of their classes, resulting in a sample that included 40 students in grade 6, 66 students in grade 7, and 34 students in grade 8. The whole middle school was actively implementing WG, having

participated in its development and having been one of the schools in an Institute-of-Education-Sciences-funded randomized controlled trial of the curriculum (see Jones et al., 2019). Analyses of the data revealed that participants tended to select either the academic language or the rich evidence essay in answer to questions about which would earn the best grade and which was most persuasive, and they responded in significantly different ways to the three topics. Because the topic-related differences were confounded by variation in the essays' formal and linguistic features, we revised the materials to remove this confound before administering the version of the FEXA we report on here.

Procedure

The revised FEXA was administered using the participating school's one-to-one Chromebooks during an assembly-style meeting of all eighth-grade classes. The two topics were programmed in a blocked design in Qualtrics so that each participant would be randomly assigned to a condition when she clicked the survey link, which was posted by a school administrator. Students were given 30 minutes to complete the task and could proceed through it at their own pace.

One teacher read aloud a paraphrase of the instructions that were printed on the first page of the Qualtrics survey. The students logged into the survey with their school-issued student identification codes, at which point they were randomly assigned to either the Recycling or the Uniform condition. Then the four essays were presented one at a time on the screen in random order; the student would click a button to advance from one screen to the next, and the timing of these clicks was recorded for later analysis. The questions were presented in a fixed order with the four essays presented alongside them for reviewing purposes.

The FEXA was adaptive to students' responses in its solicitation of their explanation of choosing for the Grade and Friend questions the same or two different essays. If a student selected the same essay, she received a prompt that repeated the name of the selected essay's author and acknowledged the sameness of those choices before asking for an explanation. If she had selected two different essays for those questions, her next prompt would remind her of her choices before asking for an explanation. Finally, after the participant made her choice of least persuasive essay in response to the last question, Qualtrics would redirect to a page displaying only that essay (as an image, so as to preclude copying and pasting), with an adjacent text box for typing the revision.

Analytic plan

Multiple-choice responses were analyzed using multiple comparisons Tukey contrasts and Fisher's exact tests to account for the categorical nature of the data. Where the focus was on the prediction of a binary response, regression analyses were used, and the predictive value of demographic variables and Partnership for Assessment of Readiness for College and Careers—English Language Arts (PARCC-ELA) scores was examined through their inclusion as fixed-effect terms. In addition, the explanations students gave for their selections on the Grade and Friend questions were manually coded for emergent themes related to text features and quality. To identify whether the emerging patterns would hold across the sample, we tallied the frequency of all the words used in those explanations automatically using the CLAN application (MacWhinney, 2000) and then by hand selected the words that were related to writing (e.g., "fact," "reason," "information"; see Appendix for full list). To explore the patterns of reasons that students supplied, we then ran separate frequency analyses of these writing-related words on the two focal subsets of students' explanatory responses (one containing explanations from students who had chosen the same essay in response to the Grade and Friend questions, and the other containing explanations from those who had chosen two different essays in answer to those questions).

Revised essays were analyzed by applying an adaptation of the revision coding scheme developed by Cho and MacArthur (2010), which targets three levels of revision: surface-level revisions (e.g., lexical substitutions, removing references to the first person), micro-level revisions (e.g., elaboration of existing claims or reasons), and macro-level revisions (e.g., introduction of new claims or reasons,

reorganization of paragraph elements). Given the shorter length of the revised FEXA essays relative to the essays analyzed by Cho and MacArthur, the two latter categories were collapsed into one ("deep revisions"). Two research assistants blind to the purpose of the study coded the revisions students produced after a training phase on a subset of the essays not reported on here. During the training phase, substantial agreement was reached in both categories (surface, 92%; deep, 100%). The target revised essays were randomly divided in half into two sets; a randomly selected 20% of the essays from each of the sets was also duplicated and placed into the other set to ensure continued reliability between coders. Substantial agreement on the double-coded essays was also reached in the target sample (surface, 92%; deep, 92%). Within these two large coding categories, we explored more specific features of the revisions as well, using logistic regression to examine associations between students' revisions and their selected essays in response to other questions.

Results

RQ 1: What do students' judgments reveal about the value they attach to argumentation features? Are their preferences related to their backgrounds or ELA scores?

The distribution of responses to the categorical target questions is presented in Table 2 and visualized in Figure 1. No differences between the Recycling and Uniform forms were detected in answers to any of the questions (all *p*s > .26), so we collapsed the results from both forms for all analyses.

Students drew a number of distinctions that were consistent with expectations. First, students responded differently to the Grade and Friend questions, $z = 4.08$, $p < .001$. They tended to choose the academic language essay in response to the Grade question (49% of students), while the distribution of

Table 2. Percentage of Responses (Raw Frequencies in Parentheses) to the (ELA) Grade, Friend (Interest), (Most) Persuasive, and (Least) Persuasive Questions, Collapsing Over Both Forms

	Question							
Response	Grade		Friend		+ Persuasive		– Persuasive	
Academic language	49%	(78)	25%	(40)	38%	(60)	11%	(18)
Rich evidence	28%	(45)	21%	(34)	38%	(61)	12%	(19)
Multiple perspectives	16%	(25)	20%	(32)	14%	(22)	22%	(35)
Rhetorical appeal	8%	(12)	34%	(54)	11%	(17)	55%	(88)

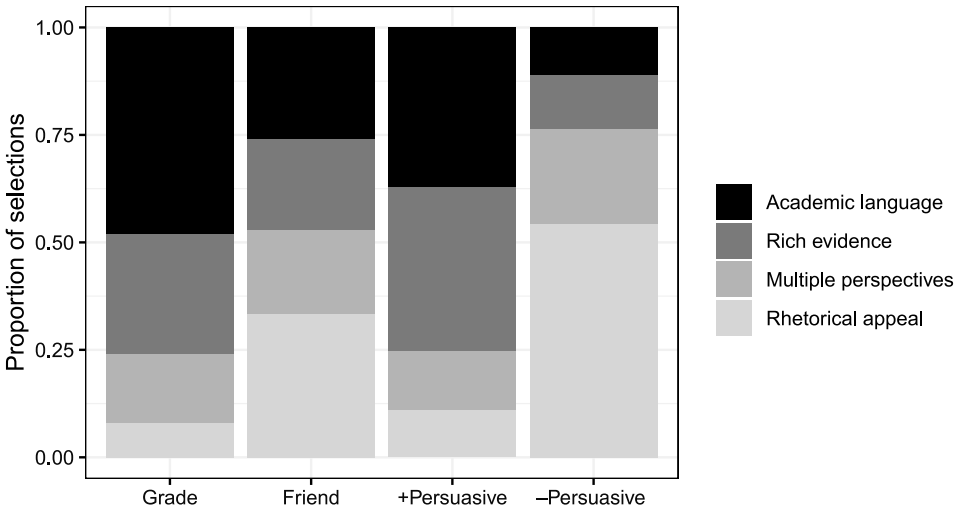

Figure 1. Distribution of responses to FEXA categorical questions, collapsing over the two topic forms.

responses to the Friend question was nearly equivalent across the four essay choices. Despite the skill in discriminating essays for different purposes that these distinct response patterns indicate, multinomial regression models revealed no significant associations between students' responses to these two questions and the demographic predictors or the standardized test score data.

Students' categorical judgments of the most and least persuasive essays also showed contrasting patterns as expected. As with the Grade question, for +Persuasive they chose the rich evidence and academic language essays frequently (at an equal rate: 38% of students chose each). For –Persuasive the predominant selection was the rhetorical appeal essay (55%). We also found significant relationships between students' responses to the –Persuasive question and their demographic characteristics, but not to their ELA scores. Overall, girls were 38% more likely than boys to choose the rhetorical appeal essay as least persuasive, $\beta = -0.96$, $p = .003$. We also found a significant difference in the selections made by students with and without an IEP, $\beta = -1.19$, $p = .03$: while only 9% of students without an IEP selected the rich evidence or academic language essay for –Persuasive, 23% of the students with an IEP did so.

RQ 2: How are beliefs about essays' purposes related to beliefs about argumentation and persuasiveness?

To answer this question we proceeded in three steps. First, we examined the distributions of students' responses to the Grade and Friend questions relative to those of their +/–Persuasive selections to identify any common patterns that would signal how students gauged persuasiveness. Second, we explored the detected contingencies across those pairs of questions according to the specific essays students chose. Third, we drilled down into the contrasts that students drew and explained around their answers to the Grade and Friend questions.

In the first set of RQ 2 analyses, then, we looked at whether students' responses to the Grade and Friend questions, separately, patterned in a way that was predictable from their selections of most and least persuasive essays (visualization in Figure 2). Using Fisher's exact tests we found that there were significant contingencies between students' responses to the Grade and +Persuasive questions, $p = .001$, as well as to the Grade and –Persuasive questions, $p = .009$. That is, students' judgments of persuasiveness differed according to their Grade responses. For the Friend question the relationship with students' –Persuasive selections was significant as well, $p < .001$, but the relationship with their +Persuasive selections was not, $p = .41$. The significant relationship between Friend and –Persuasive is remarkable, given how flat the pattern of responses to the Friend question was. It is unsurprising, in other words, that the likelihood of a student selecting, for example, the multiple perspectives essay as their +Persuasive response was no different whether they had chosen academic language, rich evidence, rhetorical appeal, or multiple perspectives on the Friend question. But the fact that students' –Persuasive responses were contingent upon their Friend selection, despite that flat distribution, suggests they may have had less consistent criteria for their –Persuasive judgment than for their +Persuasive judgment and that those criteria were in some way related to their Friend considerations.

To enrich our understanding of these patterns, we moved to the second step and examined which essays characterized the contingent responses across pairs of questions (Figure 2). In total, most students ($n = 89$) rejected as most persuasive the essay they believed would earn the highest ELA grade, and they were especially likely to do so if their Grade selection had been multiple perspectives or rhetorical appeal. Indeed, most students chose one of either academic language or rich evidence in response to *both* of these questions ($n = 100$). However, those who chose multiple perspectives for Grade nevertheless said that the academic language essay was +Persuasive more often than they did the multiple perspectives essay (45% vs. 32%), while those who chose rhetorical appeal for Grade tended to select for +Persuasive rich evidence rather than rhetorical appeal (38% vs. 11%). Next, we found that students who chose the academic language essay as their Grade response were the most consistent in choosing rhetorical appeal as the –

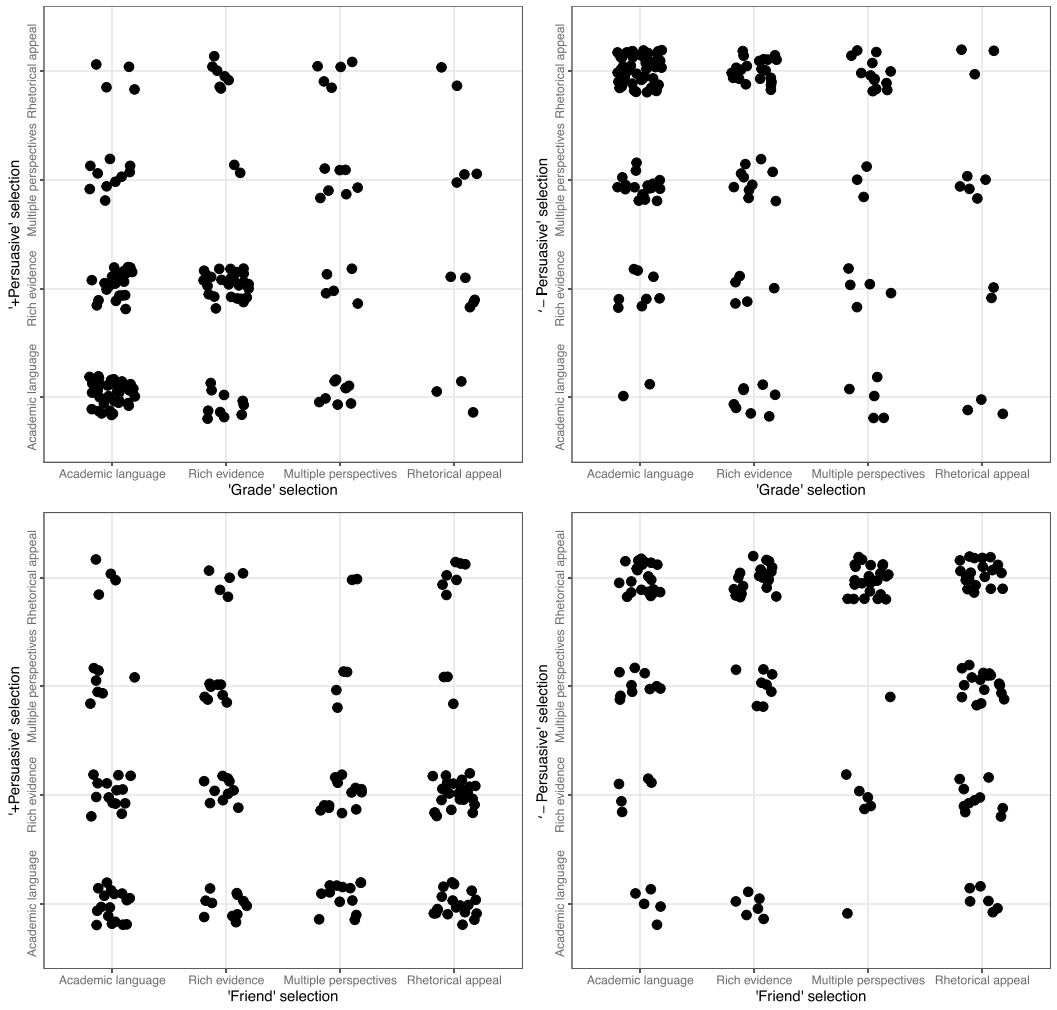

Figure 2. Distribution of responses for respective pairs of Grade/+Persuasive, Grade/–Persuasive, Friend/+Persuasive, and Friend/–Persuasive questions. These categorical selections were jittered around their points to facilitate inspection of frequencies.

Persuasive essay (64% vs. the next-highest tendency, 50%). In exploring the significant association between Friend and –Persuasive selections, we found that no student who selected the rich evidence essay as –Persuasive had chosen it for the Friend question. In addition, the tendency to choose rhetorical appeal as –Persuasive was notably stronger among those who chose multiple perspectives for their Friend (80% vs. the next-highest tendency, 59%). Overall, these analyses revealed significant variation among students' response profiles, specifically with respect to relationships between perceived usefulness for academic purposes and beliefs about the features that confer persuasiveness.

As a third step, with this range of response profiles in mind, we then tested directly whether some students were particularly sensitive to the different purposes of the Grade and Friend essay selections, irrespective of their beliefs about the persuasiveness of their choices. There was a stronger tendency in the sample for students to draw a distinction in their responses to the Grade and Friend questions ("differentiators": $n = 122$) than for them to choose the identical essay in response to both ("identifiers": $n = 38$). Students who were identifiers tended to select the academic language essay, $\beta = 1.07$, $p = .01$ (see Figure 3, specifically the rightmost block in the "Same selection" bar). That is, while the distribution of

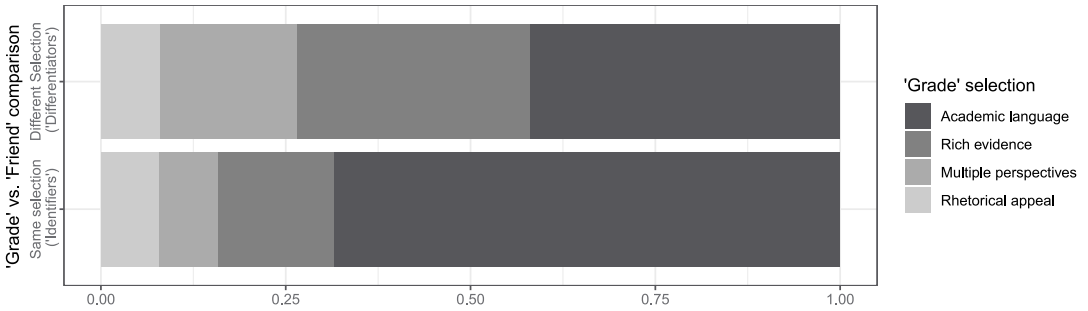

Figure 3. Distribution of responses to the Grade question (the blocks of different shades of gray in each bar) among students who selected the same (lower bar, "identifiers") and different (upper bar, "differentiators") essays for the Grade and Friend questions, scaled to unity.

responses for the Friend question was flat, choosing the academic language essay for the Grade question tended to be associated with choosing that same essay for the Friend question (26 vs. 12 total students who chose one of the three other essays in response to both questions). We can explain this result by looking at how students themselves explained their selections.

Exploring students' explanations

The Grade and Friend questions were designed to elicit evidence of a contrast in students' sensitivity to the discourse demands and varying perlocutionary force in essays characterized by different features. Representative examples of identifiers' explanations for why they had chosen the same essay for both the Grade and the Friend questions can be found in (3), where notable phrases are italicized. The examples are selected from explanations for having chosen the academic language essay, labeled as being written by Alex or Ali, and the rich evidence essay, labeled as being written by Taylor. This set of explanations tended to mention the grammar/language and register of that essay as well as the justifications for the main claim that were laid out in it.

(3) a. The passage [academic language] is *easy to read*, there is *not a lot of hard grammar and language* in it. It is a *smooth* read and *the points are valid*. It *does not use "I" in the topic sentence* and it sounds like Alex knows how to write a *decent essay*.
b. Because Ali *didn't talk in first person* and her *opening and closing statements* were better than the other 3. Ali also had the *most facts* and they were *most interesting* to read.
c. I chose Alex for both questions because his essay was *informative* and there was *no first person*. It sounded *professional* and he/she *answered the question perfectly*.
d. I think Taylor's essay [rich evidence] would receive the best grade in ELA class because it *correctly follows the essay format*. They stated their *main point* and gave at least *three supports*. Furthermore, their *supporting details* were *facts*, whereas in the other essays, there were *opinions*. Taylor's essay would be interesting for my closest friend, because the *factual information really makes the reader think about the topic*. Opinions, however, do not pull the reader in and are, overall, uninteresting.

On the other hand, differentiators, who selected a different essay in response to each of those questions, gave explanations like those reproduced in (4). The discussions here appear to recognize the more thorough and structured support that the reasons given in the academic language and rich evidence essays provided. They attributed value to evidence and academic language for the purpose of earning a high grade but found the style of the rhetorical appeal essay more interesting and entertaining.

(4) a. Sam's essay [rhetorical appeal] is much *better written* than everybody else's essay. Logan's essay [rich evidence] has *more facts* about how it would benefit the players and people buying souvenirs.
b. The first essay, Ali's [academic language] essay, had an *introduction, three points supporting* the subject, then a *conclusion*. Sam's essay [rhetorical appeal] had *more of a story* and was *more interesting to read*.
c. Taylor's essay [rich evidence] has *more facts*, and *evidence*, but Ira's [rhetorical appeal] includes more *vivid details* and in my opinion, would be *more interesting to read*.
d. Alex's story [academic language] was the best because he wrote *grammatically correct*. Ira [rhetorical appeal] and Marlo [multiple perspectives] used *first person*. Also Alex's story was slightly *better* than Taylor's [rich evidence]. Marlo's was the best to read with a friend because it was the most *entertaining*. If I was with a friend I wouldn't want to read a *boring* book I would like to read an *entertaining* book.

The automated analyses of writing-related lexical items brought out objective contrasts between the subsets of explanations written by identifiers and differentiators for their pair of responses to the Grade and Friend questions. This analysis revealed that differentiators' explanations of the contrast they drew contained an exceptionless superset of the words used in identifiers' explanations of the consistency in their responses. Several of the words used only by differentiators acknowledged structural features: "conclusion," "explanation," "paragraph(s)," and "transition(s)." Other words appearing only in differentiators' explanations referred to aspects of the genre, even though the term "persuasive" would not have been encountered in the materials until after this point: "personal," "persuasive," "supportive," "understand," "voice," and "worded."

RQ 3: How do students revise essays they find less persuasive?

Last we analyzed the revisions that students produced of the essay that they selected in response to the –Persuasive question. We focused the analysis on changes made to the most common selection for the –Persuasive question: the rhetorical appeal essay (n = 88).

Of the 88 students who made this selection, 60 supplied a revision, 12 responded with an explicit judgment about why their selected essay was least persuasive, and the remainder produced nonresponses (e.g., no comment; off-topic commentary). The 60 revisions were coded according to our adaptation of Cho and MacArthur's (2010) coding scheme. Following our operationalizations of the "surface" and "deep" categories based on that coding scheme, we explored the incidence of several key features among the revisions. The category of surface changes, detected in 93% of the revisions, included three kinds of simple repairs: deletions or simple replacements (68%), register changes at the morphosyntactic level (27%), and removal of reference to the first person (82%). The category of deep changes, detected in 98% of the revisions, included three kinds of complex repairs: addition of qualifications (63%), content extensions such as elaborating or adding claims/reasons (83%), and reorganizations such as the reordering of sentences to change the logic of the argument (53%).

Logistic regression was used to determine (1) whether students' choice of +Persuasive essay bore any relation to their revisions on their selected –Persuasive essay and (2) whether the revisions they introduced could be predicted by the essay they chose on the Grade question (i.e., whether revisions would reflect students' beliefs about academically appropriate essays). The analyses used the academic language essay as the reference category because it was the most prevalent selection in response to both the +Persuasive and Grade questions. While we did not find significant differences among the revisions students made depending on their +Persuasive selection, we did find connections between revision features and students' Grade selections. In particular, students who selected the rich evidence essay for Grade were least likely to make simple deletions or replacements, $\beta = -0.32$, $p = .02$. Additionally, students who selected the rhetorical appeal essay for Grade were most likely to incorporate register changes at the morphosyntactic level, $\beta = 0.79$, $p = .01$, when they revised that same essay later in the task. No other relationships were significant.

Discussion

The results of the first administration of the FEXA show that eighth-grade students drew sensible and desirable distinctions between the functions of differently styled and differently structured argumentative essays. They identified, as standards suggest they should, rich evidence and clear, discourse-marked language as features of high-quality essays. The findings further revealed encouraging patterns in students' explicit understandings of what makes written persuasive arguments strong, both through their meta-talk about their choices and through the choices themselves. In addition, some intriguing contrasts were documented between subgroups in their response patterns. These findings on the whole serve to illuminate students' perceptions of written arguments' suitability for different purposes and how they put those perceptions to use in the revision process. In contrast to past research that manipulated reading purposes between subjects (e.g., work discussed in Rouet et al., 2017), we reported within-participants contrasts that show notable flexibility among these adolescent students (relatedly, see Qin & Uccelli, 2016, for genre distinctions in second-language learners' writing). In effect the findings increase our understanding of what students are actually learning about argumentation pedagogy—in particular in an instructional context that provides only passing explicit guidance for this genre of writing. In this Discussion we return to our research questions, describing what these findings tell us about students' understanding of the features and functions of persuasive argumentation. We also discuss the potential for the FEXA to be used by teachers as a formative assessment of students' argumentative writing development.

Selecting essays for specific purposes

Students tended to select the academic language essay as the one most likely to get a good grade, and that essay also was selected as most persuasive just as often as the rich evidence essay was. This suggests two possibilities: that students have internalized a message of the importance of academic language in the minds of their teachers or that they genuinely believe the presence of academic-language features increases persuasiveness—something that future studies will seek to tease apart. The academic language essay variant that we presented to students was not characterized by technical jargon. Instead, we incorporated discourse markers and nominalizations, features that may be less salient as explicitly academic (Uccelli et al., 2015).

Findings from a small follow-up administration of the FEXA suggest that expert adult readers find the academic language essay likely to get the best grade and the rich evidence essay the most persuasive, like a large number of our participants did. This correspondence, should it stand up to more systematic data collection with adults, points to a maturity in our student participants' judgments. Existing developmental research on judgments of argumentativeness suggests that the specific characterizations of texts and their purposes do influence people's judgments differently as they get older (Golder & Coirier, 1996). Future research with participants from an extended range of ages will therefore seek to use the FEXA to detect nuanced shifts in judgments during development.

In contrast to responses to the Grade, +Persuasive, and –Persuasive questions, the four essay types were all equally likely to be chosen as most interesting for a friend to read, indicating individual differences in preferences that the other questions did not elicit. The essay students found most interesting, in other words, was not necessarily the essay that they found most persuasive (or that they believed would earn the best grade). If interest and persuasiveness do not show much overlap, this may be one source of the difficulty of maintaining civility in argument more generally.

Explaining identified contrasts

The explanations that students gave for choosing the same or a different essay for the Grade and Friend questions revealed a notable amount of attention to the essays' structure and content.

Students who were *identifiers* justified their selection of the academic language essay for both of those purposes on grounds of readability. These students particularly remarked the register of the academic language essay, observing its "professional" sound and its lack of use of the first person. On the other hand, the *differentiators* who chose distinct essays in answer to these two questions reported not finding the academic language essay as interesting as some of the other options. The rhetorical appeal essay told "more of a story," while the rich evidence essay contained "more facts." None of the collected demographic or achievement data explained students' varying response patterns on this pair of items, but the mere existence of those patterns suggests that some students are more sensitive than others to the differing demands of persuasive writing, for different purposes.

These patterns are reminiscent of detectable differences in students' argumentative writing production in which a general goal of "persuasion" leads to weaker essays than the adoption of subgoals of discrete elements of argumentative discourse (see Ferretti et al., 2009). The finding that even momentarily considering the purpose of an essay, as students were asked to do in the FEXA, elicited evidence of different attitudes and beliefs about argumentative texts, suggests that students' sensitivity to texts' features—beyond their argumentative structure—could be fostered through instruction (see, e.g., Kuhn et al., 2016; Reznitskaya et al., 2001).

Accounting for differences in essay selections

The patterns of responses to individual questions on the FEXA were in some cases significantly related to students' demographic characteristics. Boys did not find the rhetorical appeal essay as unpersuasive as girls did, for example. If we assume that the FEXA is drawing out some of the same skills that students apply in their own writing, then this finding seems consistent with teachers' ratings of girls as better writers (e.g., Pajares et al., 2007). That is, if girls are more apt to recognize that the features embodied by the rhetorical appeal essay make for less persuasive writing, then we could also expect that they would use those features less often in their own persuasive writing—which in turn would lead to teachers rating girls' writing as being of higher quality. Additionally, students with IEPs selected the essays with rich evidence and academic language in answer to the –Persuasive question more often than students without IEPs. It appears that these students would benefit from additional support in strategies to recognize markers of bad arguments as well as of good ones, well beyond the common five-paragraph structure (see Brannon et al., 2008). Because we were unable to examine the sample of students with IEPs for any detail about their exceptionalities, we can only hazard conjectures about why many of them found the rich evidence and academic language essays least persuasive—having to do, perhaps, with their more challenging vocabulary and morphosyntax or with their less event-based and more abstract content. Finally, we did not detect any relationship between performance on the Common-Core-aligned PARCC-ELA assessment and the FEXA responses. The measures appear to be tapping into different skills, and perhaps the metacognitive aspects of the FEXA are not recruited during engagement with the PARCC, despite its presentation of a range of question and response types.

The significant relationships we found between the types of revisions that students applied and their selections on the Grade question suggested a degree of systematicity in their interactions with the texts. On one hand, we found that students who believed the rich evidence essay would earn the highest grade in ELA class made the fewest simple deletions or replacements when revising the rhetorical appeal essay. We interpret this finding to suggest that students who identified rich evidence as the most important feature for an ELA submission may persist in that thematic focus even when making revisions. That is, the structure of the essay may have been less salient to them than the substance was. On the other hand, we also discovered that students who believed the rhetorical appeal essay was least persuasive but who had also asserted it would earn the highest ELA grade were likeliest to make morphosyntactic changes that raised the essay's register. Hedging any interpretation against the small number of students in this group, this result recalls the literature on revisions that reports that students tend to make superficial changes when revising their writing (as morphosyntactic

changes in a sense are). That is, given the stated purpose of the revision as increasing persuasiveness, students who thought rhetorical appeal would earn the highest ELA grade—in contrast to the vast majority of their peers—may have opted to take a relatively simple, if still potentially effective, route in their revision efforts.

Revising essays judged least persuasive

Some of the features that students cited as making them choose the same essay for the Grade and Friend questions were also reflected in the changes they introduced in their revisions. Many students focused on superficial features of academic language, such as exchanging words that can be perceived as lower-register words for higher-register ones (e.g., "receive" instead of "get") or changing the grammatical persons mentioned in the essay from a combination of first, second, and third person to third person only. While superficial, these changes indicate that students have grasped some version of the notion of linguistic register in academic writing but may not yet have identified the relationships between specific language forms and persuasiveness (see Troyer, 2017).

However, the more substantive changes that some students made were apparently intended to increase persuasiveness, such as elaborating or adding reasons to support an essay's main claim, or reorganizing the paragraph. They also added hedges and qualifications to the reasons that were provided in the original essay, demonstrating a mature understanding of what it means for an essay to be persuasive. These changes are somewhat more in line with the broader characterization of persuasiveness promoted by Cialdini and others. According to that perspective, awareness of an audience and the beliefs (and skepticism) a reader might bring to an argument can play a role in shaping how that argument should be presented. It appears that students believe—rightly—that persuasiveness is often served by combining the precision of academic language with the careful elaboration of reasons. To address this possibility, in future versions of the task we will ask students to rank the essays rather than selecting a single best answer, probe their choices by eliciting more explanations, and apply subtler manipulations of essay features.

Practical implications of the FEXA

In addition to informing our understanding of students' capacities to judge and revise argumentative writing, the development and use of the FEXA has lessons for instructional practice. Designing the FEXA iteratively and with the input of teachers and students produced a measure that teachers can use and that students enjoy. Teachers and administrators expressed appreciation for what the measure revealed about their students' beliefs and preferences, according to informal conversations carried out during the measure's development. They foresaw using what they learned from a summary of findings to inform lessons on argumentative writing, as an aid to determining the level of detail they should provide about language features, quantity of evidence, and discourse markers.

Teachers also suggested they could use the FEXA as a way to sensitize students to different writing purposes and stylistic options. Students in both the pilot and target studies sometimes gave opinions on the task in addition to (and occasionally instead of) supplying responses. It was clear that they were engaged with the measure and enthusiastic about having their academic judgments elicited. Considering students' willingness to reflect on writing styles and justify their resulting articulated beliefs, the FEXA could also be thought of as a step toward or a component of instruction in cognitive strategies for writing or in metacognition more generally.

In particular, the variation in students' selections suggests there may be untapped potential in students' sensitivity to varying features of written persuasive argument. The FEXA could help teachers draw out that potential through several possible uses in classrooms. First, eliciting the judgments contained in the forced-choice questions could help students reflect on what writing features are valued in school and how those overlap with or diverge from the values of their own community of students, especially in an age of so much (digital) written communication among adolescents. Making choices and then trying to explain

them could make them more aware of the features their own writing should display when set to distinct purposes. Teachers could use the FEXA to help students see that writing for an audience of peers as opposed to authority figures may call for different argumentation features but that audience identification is just one of many factors that can influence the shape and contents an argument should have. Teachers could also adapt the measure to their own goals, holding the essay sets constant but adding further questions about judgments and purposes that align to their pedagogical needs.

Second, instructions for revising the least persuasive essay were not to "fix" that essay but to "make points . . . in a better way," allowing for stylistic as well as substantive revisions. This low-stakes task of revising someone else's work—and not even a peer's but a hypothetical student's, and not for a grade—could be useful in clarifying for students the several ways in which writing can be improved through revision, especially with a purpose (or an audience, etc.) in mind. The "safe" FEXA peer revision process could be used as a transition toward instituting a robust and regular peer assessment practice, with its attendant academic advantages.

Third, the FEXA's selection, explanation, and revision exercises could be used by teachers as a formative assessment of the degree of nuance that their students are developing in their understanding of written persuasive argumentation. Together these exercises could serve as a springboard for more customized lessons that elaborate the variety of writing styles and structures that can succeed in persuading. We plan to explore these possibilities in the future by collecting pre- and post-writing intervention data from students and by eliciting feedback from their teachers.

Overall, the findings from the administration of the FEXA reported here suggest key similarities and contrasts with adult expert readers' approaches to the task and to lessons from the literature on argumentation. The student respondents appeared to focus on superficial features in their revisions, in line with the Cho and MacArthur (2010) finding that single-reader revisions emphasize surface-level improvements. However, in contrast to that work, FEXA participants also frequently made deeper revisions involving new or elaborated reasons. The brevity of the essays they were asked to revise made it unlikely that their writing would realize a complete argument schema (as in, e.g., Reznitskaya et al., 2001), but they did tend to increase the support for the position taken in their arguments.

The range of response patterns we found across the multiple-choice questions reflected the varying skills that reading experts teach, or believe need to be taught, in this genre as well (e.g., Sundeen, 2015; see also NGAC, 2010). Many students chose the rich evidence or the academic language essays as likely to get the best Grade and being most Persuasive, aligning with teacher-focused literatures on argumentative writing (e.g., Dougherty Stahl, 2014) and the academic register (e.g., Crosson et al., 2012; Uccelli et al., 2013).

Limitations

Despite the suggestiveness of these findings, the study is limited in a variety of ways. Due to its exploratory nature, the relationships we observed and the claims we have made are merely suggestive of hypotheses that we will test in future work. We are particularly interested in further understanding the relationship between students' expressed *beliefs* about argumentation in the FEXA and related assessments and their *ability* to detect and produce the relevant features in various contexts. Among the more practical study limitations, the sample was drawn from a relatively homogeneous district, so we were unable to explore the range of potentially relevant demographic differences reflected in the larger U.S. student population. The participants also each completed only one form, so while analyses indicated there were no meaningful differences between the forms, we are unable to speak to test-retest reliability across the forms or within individuals. Because all participants were being exposed to WG, the results may not generalize to populations of students who have not experienced regular debates on contentious topics in their classrooms (although argumentative *writing* instruction in WG was sparse). In future studies we intend to validate the measure by incorporating a holistic measure of persuasiveness (e.g., Ferretti et al., 2009) and explore its utility with different populations of students, such as students in urban contexts, English learners, and students in reading remediation programs.

Conclusion

Mastery of written argumentation is critical to academic success and to responsible participation in citizenship. This is an era of ever more communication through electronic media, in which virtually anyone can share an opinion in a text. It falls to educators and education researchers to teach the next generation how to mold content worth knowing into ideas worth having and to compose these into arguments in coherent, convincing ways. Bringing all students into the practices of argumentation, both producing arguments and gauging the quality and persuasiveness of others' arguments, is yet another way to take the best of our social practices and support students in internalizing them, so they can seize every available educational opportunity, formal and informal alike. Measures like the FEXA that help researchers and educators see where students stand with respect to those practices are a useful tool for the pursuit of that end.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Lisa B. Hsin http://orcid.org/0000-0003-0414-9964

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*(2), 135–167. https://doi.org/10.1207/s1532690xci1502_1

Andrews, R., Torgerson, C., Low, G., & McGuinn, N. (2009). Teaching argument writing to 7-to 14-year-olds: An international review of the evidence of successful practice. *Cambridge Journal of Education*, *39*(3), 291–310. https://doi.org/10.1080/03057640903103751

Basturkmen, H., & von Randow, J. (2014). Guiding the reader (or not) to re-create coherence: Observations on postgraduate student writing in an academic argumentative writing task. *Journal of English for Academic Purposes*, *16*, 14–22. https://doi.org/10.1016/j.jeap.2014.07.005

Brannon, L., Courtney, J. P., Urbanski, C. P., Woodward, S. V., Reynolds, J. M., Iannone, A. E., Haag, K. D., Mach, K., Manship, L. A., & Kendrick, M. (2008). EJ Extra: The five-paragraph essay and the deficit model of education. *The English Journal*, *98*(2), 16–21.

Carrell, P. L. (1995). The effect of writers' personalities and raters' personalities on the holistic evaluation of writing. *Assessing Writing*, *2*(2), 153–190. https://doi.org/10.1016/1075-2935(95)90011-X

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*(4), 328–338. https://doi.org/10.1016/j.Learning and Instruction.2009

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, *98*(4), 891–901. https://doi.org/10.1037/0022-0663.98.4.891

Cialdini, R. B. (2004). The science of persuasion. *Scientific American Mind*, *14*(1), 70–77. http://www.jstor.org/stable/24939368

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*(1), 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Clark, R. A., & Delia, J. G. (1976). The development of functional persuasive skills in childhood and early adolescence. *Child Development*, *47*(4), 1008–1014. https://doi.org/10.2307/1128437

Crossley, S. A., Roscoe, R., & McNamara, D. S. (2011). Predicting human scores of essay quality using computational indices of linguistic and textual features. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial intelligence in education* (pp. 438–440). Springer Berlin Heidelberg.

Crossley, S. A., Roscoe, R., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication*, *31*(2), 184–214. https://doi.org/10.1177/0741088314526354

Crosson, A. C., Matsumura, L. C., Correnti, R., & Arlotta-Guerrero, A. (2012). The quality of writing tasks and students' use of academic language in Spanish. *The Elementary School Journal*, *112*(3), 469–496. https://doi.org/10.1086/663298

Crowhurst, M. (1990). Teaching and learning the writing of persuasive/argumentative discourse. *Canadian Journal of Education / Revue canadienne de l'éducation*, *15*(4), 348. https://doi.org/10.2307/1495109
Cutler, L., & Graham, S. (2008). Primary grade writing instruction: A national survey. *Journal of Educational Psychology*, *100*(4), 907–919. https://doi.org/10.1037/a0012656
Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*(1), 7–24. https://doi.org/10.1016/j.asw.2012.10.002
Deane, P., Song, Y., van Rijn, P., O'Reilly, T., Fowles, M., Bennett, R., & Zhang, M. (2018). The case for scenario-based assessment of written argumentation. *Reading and Writing, 32*(6), 1575-1606.
Dougherty Stahl, K. A. (2014). What counts as evidence? *The Reading Teacher*, *68*(2), 103–106. https://doi.org/10.1002/trtr.1318
Duran Gisbert, D., & Monereo Font, C. (2008). The impact of peer tutoring on the improvement of linguistic competence, self-concept as a writer and pedagogical satisfaction. *School Psychology International*, *29*(4), 481–499. https://doi.org/10.1177/0143034308096437
Early, J. S., & Saidy, C. (2014). A study of a multiple component feedback approach to substantive revision for secondary ELL and multilingual writers. *Reading and Writing*, *27*(6), 995–1014. https://doi.org/10.1007/s11145-013-9483-y
Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, *32*(4), 400–414. https://doi.org/10.2307/356602
Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skill. *Discourse Processes*, *32*(2–3), 135–153. doi:10.1080/0163853X.2001.9651595
Felton, M. K. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development*, *19*(1), 35–52. https://doi.org/10.1016/j.cogdev.2003.09.001
Ferretti, R. P., Lewis, W. E., & Andrews-Weckerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology*, *101*(3), 577–589. https://doi.org/10.1037/a0014702
Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology*, *92*(4), 694. https://doi.org/10.1037/0022-0663.92.4.694
Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research*, *57*(4), 481–506. https://doi.org/10.3102/00346543057004481
Fitzgerald, J., & Markham, L. R. (1987). Teaching children about revision in writing. *Cognition and Instruction*, *4*(1), 3–24. https://doi.org/10.1207/s1532690xci0401_1
Gelfert, A. (2011). Expertise, argumentation, and the end of inquiry. *Argumentation*, *25*(3), 297. https://doi.org/10.1007/s10503-011-9218-7
Golder, C., & Coirier, P. (1996). The production and recognition of typological argumentative text markers. *Argumentation*, *10*(2), 271-282. https://doi.org/10.1007/BF00180729
Graff, G. (2003). *Clueless in academe: How obscuring obscures the life of the mind*. Yale University Press.
Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*(3), 445–476. https://doi.org/10.1037/0022-0663.99.3.445
Hoogeveen, M., & van Gelderen, A. (2013). What works in writing with peer response? A review of intervention studies with children and adolescents. *Educational Psychology Review*, *25*(4), 473–502. https://doi.org/10.1007/s10648-013-9229-z
Hsin, L. B., Phillips Galloway, E., and Snow, C. E. (under review). Supporting social strengths amid emerging bilingualism: Sociocognitive effects of Word Generation in English-language learners' argumentative writing.
Jones, S. M., LaRusso, M., Kim, J., Kim, H. Y., Selman, R., Uccelli, P., & Snow, C. (2019). Experimental effects of word generation on vocabulary, academic language, perspective taking, and reading comprehension in high poverty schools. *Journal of Research on Educational Effectiveness, 12*(3), 448–483. https://doi.org/10.1080/19345747.2019.1615155.
Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science*, *39*(3), 387–406. https://doi.org/10.1007/s11251-010-9133-6
Kuhn, D., & Crowell, A. (2011). Dialogic Argumentation as a Vehicle for Developing Young Adolescents' Thinking. *Psychological Science*, *22*(4), 545–552. https://doi.org/10.1177/0956797611402512
Kuhn, D., Cummings, A., & Youmans, M. (2020). Is reasoning a fruitful path to changing minds? *Discourse Processes*, *57*(1), 36–47. https://doi.org/10.1080/0163853X.2019.1646566
Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how)? In W. Damon, R. Lerner, D. Kuhn, & R. Siegler (Eds.), *Handbook of child psychology*, *Cognition, perception and language* (6th ed., vol. 2, pp. 953– 993). John Wiley.
Kuhn, D., Hemberger, L., & Khait, V. (2016). Tracing the development of argumentive writing in a discourse-rich context. *Written Communication*, *33*(1), 92–121. https://doi.org/10.1177/0741088315617157
Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking & Reasoning*, *13*(2), 90–104. https://doi.org/10.1080/13546780600625447
Kuhn, D., Wang, Y., & Li, H. (2010). Why argue? Developing understanding of the purposes and values of argumentive discourse. *Discourse Processes*, *48*(1), 26–49. https://doi.org/10.1080/01638531003653344
Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentive competence. *Cognition and Instruction*, *31*(4), 456–496. https://doi.org/10.1080/07370008.2013.830618

Lapp, D., & Fisher, D. (2012). Persuasion = Stating and arguing claims well. *Journal of Adolescent & Adult Literacy*, *55* (7), 641–644. https://doi.org/10.1002/JAAL.00076

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

Maloney, J., & Simon, S. (2006). Mapping children's discussions of evidence in science to assess collaboration and argumentation. *International Journal of Science Education*, *28*(15), 1817–1841. https://doi.org/10.1080/09500690600855419

Mayweg-Paus, E., Macagno, F., & Kuhn, D. (2016). Developing Argumentation Strategies in Electronic Dialogs: Is Modeling Effective? *Discourse Processes*, *53*(4), 280–297. https://doi.org/10.1080/0163853X.2016.1040323

McLuhan, M. (2009). *The classical trivium: The place of thomas nashe in the learning of his time* (W. T. Gordon, ed., Reprint ed.). Gingko Press Inc.

Melzer, D. (2009). Writing assignments across the curriculum: A national study of college writing. *College Composition and Communication*, *61*(2), W240.

Mendonça, C. O., Johnson, K. E., & Mendonca, C. O. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, *28*(4), 745–769. https://doi.org/10.2307/3587558

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*(1–2), 131–151. https://doi.org/10.1007/s11145-007-9067-9

Min, H.-T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, *15*(2), 118–141. https://doi.org/10.1016/j.jslw.2006.01.003

National Governors Association Center for Best Practices & Council of Chief State School Officers, & others. (2010). *Common core state standards for English language arts and literacy in history/social studies, science, and technical subjects*. CCSSO Council of Chief State School Officers.

Nussbaum, E. M. (2005). The effect of goal instructions and need for cognition on interactive argumentation. *Contemporary Educational Psychology*, *30*(3), 286–313. https://doi.org/10.1016/j.cedpsych.2004.11.002

Olinghouse, N. G., & Wilson, J. (2012). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, *26*(1), 45–65. https://doi.org/10.1007/s11145-012-9392-5

Pajares, F., Johnson, M. J., & Usher, E. L. (2007). Sources of Writing Self-Efficacy Beliefs of Elementary, Middle, and High School Students. *Research in the Teaching of English*, *42*(1), 104–120.

Perelman, C. (2012). *The new rhetoric and the humanities: Essays on rhetoric and its applications*. Springer Science & Business Media.

Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing*, *33*, 3–17. https://doi.org/10.1016/j.jslw.2016.06.001

Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes*, *32*(2–3), 155–175. doi:10.1080/0163853X.2001.9651596

Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist*, *52*(3), 200–215. https://doi.org/10.1080/00461520.2017.1329015

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics*, Vol. 2. Reading, Writing, and Language Learning (pp. 142–175). Cambridge University Press.

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, *2*(4), 325–344. https://doi.org/10.1080/19345740903167042

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). Cambridge: Cambridge University Press.

Spycher, P. (2007). Academic writing of adolescent English learners: Learning to use "although.". *Journal of Second Language Writing*, *16*(4), 238–254. https://doi.org/10.1016/j.jslw.2007.07.001

Strachan, I. B., & Wilcox, S. (1996). Peer and self assessment of group work: Developing an effective response to increased enrolment in a third-year course in microclimatology. *Journal of Geography in Higher Education*, *20*(3), 343–353. https://doi.org/10.1080/03098269608709377

Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, *20*(4), 291–303. https://doi.org/10.1016/j.learninstruc.2009.08.008

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, *20*(4), 265–269. https://doi.org/10.1016/j.learninstruc.2009.08.002

Sundeen, T. H. (2015). Writing instruction for adolescents in the shadow of the common core state standards. *Journal of Adolescent & Adult Literacy*, *59*(2), 197–206. https://doi.org/10.1002/jaal.444

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*(4), 339–343. https://doi.org/10.1016/j.learninstruc.2009.08.003

Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.

Troyer, M. (2017). 'This is what I know:' Use of the first person in sixth grade argumentative writing. *Linguistics and Education*, *38*, 24–32. https://doi.org/10.1016/j.linged.2017.02.002

Tsagari, D., & Meletiadou, E. (2015). Peer assessment of adolescent learners' writing performance. *Writing & Pedagogy*, *7* (2–3), 305–328. https://doi.org/10.1558/wap.v7i2-3.26457

Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., & Sánchez, E. (2015). Core academic language skills: An expanded operational construct and a novel instrument to chart school-relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics*, *36*(5), 1077–1109. https://doi.org/10.1017/S014271641400006X

Uccelli, P., Dobbs, C. L., & Scott, J. (2013). Mastering academic language: Organization and stance in the persuasive writing of high school students. *Written Communication*, *30*(1), 36–62. https://doi.org/10.1177/0741088312469013

Villamil, O. S., & Guerrero, M. C. M. D. (1998). Assessing the impact of peer revision on L2 writing. *Applied Linguistics*, *19*(4), 491–514. https://doi.org/10.1093/applin/19.4.491

Voss, J. F., & Dyke, J. A. V. (2002). Argumentation in Psychology: Background Comments. *Discourse Processes,32*(2–3), 89–111. https://doi.org/10.1080/0163853X.2001.9651593

Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication*, *26*(2), 183–209. https://doi.org/10.1177/0741088309333019

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology*, *94*(4), 660–668. https://doi.org/10.1037/0022-0663.94.4.660

Zohar, A., & Nemet, F. (2001). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, *39*(1), 35–62. doi:10.1002/(ISSN)1098-2736

Appendix. Full list of writing-related words appearing in students' explanations for comparison between Grade and Friend question responses. Asterisks indicate wildcard terms (i.e., including several possible strings/morphemes).

agree	easier	important	point*	support*
answer*	easy	information	positive	thinking
back*	evidence	informative	professional	topic
based	experience	interesting	question*	transition*
better	explain*	language	reader*	understand*
boring	explanation	multiple	reason*	vocabulary
clear	fact*	opening	sense	voice
conclusion	factors	opinion*	sentence*	word*
conversation	factual	paragraph*	show*	writing
correct	formal	passage	state*	
descriptive	good	personal	strong	
detail*	ideas	persuasive	studies	