

Classroom Writing Tasks and Students' Analytic Text-Based Writing

Lindsay Clare Matsumura

Richard Correnti

Elaine Wang

*University of Pittsburgh, Pennsylvania,
USA*

ABSTRACT

The Common Core State Standards emphasize students writing analytically in response to texts. Questions remain about the nature of instruction that develops students' text-based writing skills. In the present study, we examined the role that writing task quality plays in students' mastery of analytic text-based writing. Text-based writing tasks ($N = 149$) were collected from 27 fifth-grade teachers in an urban district, and teachers completed daily surveys (i.e., instructional logs) to assess the frequency of their reading and writing instruction (30-45 days total). Students ($N = 793$) completed a performance assessment of their text-based writing skills. Results showed that the large majority of writing tasks guided students to retrieve isolated facts or engage with surface-level features of texts in their writing (i.e., were of a low level of cognitive demand). The cognitive demand of text-based writing assignments predicted multiple features of students' writing performance, including students' ability to reason analytically about texts (effect size [ES] = .46), use evidence to support their claims (ES = .46), and organize their writing (ES = .35), even after controlling for other dimensions of literacy instruction. The quality (grist) of text to which students responded predicted one dimension of students' writing performance, use of evidence to support their claims (ES = .37). Designing cognitively demanding text-based tasks should be considered an essential part of writing instruction reform and professional development programs for teachers that aim to increase students' writing skills aligned to the Common Core.

Writing is critical to academic and economic success (National Commission on Writing in America's Schools and Colleges [NCW], 2003). However, results of the National Assessment of Education Progress indicate that only about a quarter of students meet standards for proficient writing (National Center for Education Statistics, 2012). Students who cannot write well are at a significant disadvantage. Beginning in the upper elementary grades, writing is a critical tool for developing thinking skills and subject matter content knowledge, as well as for expressing what one knows (Bangert-Drowns, Hurley, & Wilkinson, 2004; Hillocks, 1984, 2006). Adults who cannot write well often find college course work too difficult to complete, and face significant difficulty advancing in the workplace (Business Roundtable, 2009).

That students struggle with writing is unsurprising, given that writing has long been a neglected area of the curriculum (NCW, 2003). Research shows that students are rarely asked to write in ways that require higher level thinking (Applebee & Langer, 2011; Gilbert & Graham, 2010). Gilbert and Graham found that students in grades

4–6 spent an average of only 25 minutes a day writing and that the most common writing activities in which students engaged (e.g., short responses to questions, worksheets, note-taking, making lists) involved little to no opportunity to think analytically or produce an extended response. Even in the secondary grades, students' exposure to writing remains similarly impoverished (Applebee & Langer, 2011).

In response to public concern, the Common Core State Standards emphasize writing, especially students' ability to write analytically in response to both fiction and nonfiction (informational) texts (National Governors Association Center for Best Practices & Council of Chief State School Officers [NGA Center & CCSSO], 2010). To put students on a track to college readiness, upper elementary–grade students are expected to write extended essays in which they draw on multiple text sources to support analysis and opinion. Students must reason and construct knowledge from texts in several different ways, including comparing themes across texts, analyzing conflicting points of view, and showing how authors use reasons and evidence to support a particular viewpoint (NGA Center & CCSSO, 2010). The Standards also emphasize the importance of students stating clear claims and supporting assertions with appropriate text-based reasons and information.

Although the Common Core defines a vision of the text-based analytic writing skills that students are to learn, the path for developing these competencies is less clear. Questions remain about the forms of instruction that best develop students' analytic text-based writing skills. These questions must be answered because they inform interventions that we would develop to improve writing instruction. Reform policy recommendations emphasize the importance of increasing the amount of time students spend engaged in classroom writing (NCW, 2003). Prior research and recommendations for practitioners also emphasize process-oriented approaches (e.g., extended opportunities for writing, engaging in cycles of planning, translating and revising) and explicit attention to building students' metacognition throughout the writing process (e.g., planning and goal setting, accessing prior knowledge, self-monitoring; Applebee & Langer, 2013; S. Graham & Harris, 1993, 2006; S. Graham & Perin, 2007; Harris & Graham, 2009). Studies show that direct instruction in constructing more complex sentences, the study of text models, attention to word processing, and the use of writing to support learning across the curriculum advance students' writing development (see studies reviewed in S. Graham & Perin, 2007). All of these features of effective writing instruction, however, would not on their own constitute a complete writing program. To achieve the ambitious Common Core learning goals, students require instruction that focuses on analytic text-based writing.

The aim of this study is to contribute to our understanding of instruction that supports writing competencies aligned to the Common Core. Specifically, we examine the role that writing tasks play in the development of students' analytic text-based writing skills, while controlling for other dimensions of literacy instruction. Our goal is to generate information that contributes to creating more effective writing programs and professional development (PD) interventions for teachers.

Academic Tasks and Learning

Our focus on tasks is based on research across disciplines that shows that the cognitive demand of the tasks that teachers assign to students is critical to students' learning. Students' engagement with tasks that guide them to process and manipulate information in complex ways has been associated with increased student achievement across content areas, including mathematics (Stein & Lane, 1996), English language arts (Langer & Applebee, 1987; Marshall, 1987), and history (Monte-Sano & De La Paz, 2012; Voss & Wiley, 1997; Wiley & Voss, 1999). The kind of writing that students are asked to undertake in school directly influences student thinking (Applebee, 1984; Langer & Applebee, 1987; Newkirk & Atwell, 1982). Wiley and Voss found that students' understanding of historical subject matter increased when they were guided to write arguments instead of narratives, summaries, or explanations. Monte-Sano and De La Paz (2012) showed that tasks that led high school students to consider author bias, corroborate evidence, and explore the context and relationships among historical events increased their historical reasoning skills.

In addition to influencing students' thinking and subject matter content knowledge, research shows that the cognitive demand of the writing tasks assigned to students contributes to the development of their reading skills (Carbonaro & Gamoran, 2002; Correnti, Matsumura, Hamilton, & Wang, 2012; S. Graham & Hebert, 2010; Matsumura, Garnier, Pascal, & Valdés, 2002; Matsumura, Garnier, Slater, & Boston, 2008; Newmann, Bryk, & Nagaoka, 2001). Carbonaro and Gamoran found that high school students with more opportunities to engage in analytical writing showed higher gains in reading achievement test scores. Newmann and colleagues found that students in grades 3, 6, and 8 who were exposed to tasks that demanded interpretation, analysis, and evaluation skills showed greater gains in reading skills, compared with those exposed to less demanding writing tasks, controlling for students' prior achievement and demographics. Urban middle school students who had access to high-quality writing tasks also showed greater gains on standardized tests of reading skills (i.e., the SAT-10) compared with

peers who were exposed to less cognitively demanding writing tasks (Matsumura et al., 2008).

Finally, some research suggests that opportunity to engage with cognitively demanding writing tasks can improve the quality of students' writing (Boscolo & Carotti, 2003; Crosson, Matsumura, Correnti, & Arlotta-Guerrero, 2012; Matsumura, Patthey-Chavez, Valdés, & Garnier, 2002). Boscolo and Carotti, for example, found that high school students who engaged with analytic writing tasks generated higher quality commentaries on literary texts than students who participated in traditional class activities did. In the elementary grades, cognitively demanding assignments that encourage elaborated responses also have been associated with increased quality of the content of students' writing (Matsumura, Patthey-Chavez, et al., 2002) and use of most features of academic language in their writing (e.g., academic vocabulary, embedded clauses, temporal and causal connectives, a variety of connectives) for the student work associated with those tasks (Crosson et al., 2012). Very little research, however, has examined the relationship between the quality of text-based writing tasks and students' analytic text-based writing skills assessed on an independent writing measure, or considered the relationship of tasks and multiple features of students' text-based writing skills.

Framework for Conceptualizing Text-Based Writing Assignment Quality: Cognitive Demand and Grist of Text

Our conceptualization of the quality of text-based writing assignments is grounded in Doyle's (1983, 1988; Doyle & Carter, 1984) theoretical framework for thinking about academic tasks across disciplines. He argued that academic tasks provide a structure for students' learning and understanding of a curricular domain because they draw students' attention to particular content and guide the cognitive operations needed to process that content. From Doyle's perspective, tasks are critical to how students experience a discipline and come to understand the meaning of a curricular domain (e.g., what it means to write in response to a text or engage in mathematics). He identified four types of academic tasks, representing the categories of underlying cognitive operations that shape student learning: memory tasks, procedural or routine tasks, comprehension or understanding tasks, and opinion tasks.

Memory tasks direct students to recognize or reproduce information that they have seen before (e.g., recall a list of terms). Procedural or routine tasks require students to apply a predictable formula (e.g., solve a set of

multiplication problems) or process (e.g., a search-and-match strategy to locate words or passages in a text). Comprehension or understanding tasks, in contrast, guide students to apply or choose among procedures to solve a problem, or to draw inferences from previously encountered information or procedures (e.g., make a prediction based on past actions of a character or events). Finally, in opinion tasks, students are expected to state a preference for something. A fundamental distinction between memory and comprehension tasks is that the former tend to focus on surface-level features, whereas the latter focus on the conceptual structure and the meaning of texts (i.e., words, sentences).

Because they are part of an evaluation system, Doyle (1983) further argued that academic tasks occur under conditions of ambiguity and risk for students. *Ambiguity* refers to the extent to which a correct answer can be defined in advance. Comprehension and opinion tasks generally do not have single correct answers. They are inherently more ambiguous than memory or procedural tasks. Memory and routine tasks, in contrast, lack much of this ambiguity because they require basic reproduction of previously encountered information or the application of a set procedure.

Risk refers to the standards that teachers apply for completed work. If any answer is acceptable in the case of an unsupported opinion, for example, then the task would be considered low risk. Tasks that have a more defined range of acceptable answers would be considered to carry higher risk. Deliberate ambiguity and risk are desirable within Doyle's (1983) framework.

Our lens for examining task quality draws on past literacy research. We consider higher quality (i.e., more cognitively demanding) tasks that guide students to construct, rather than reproduce, knowledge from texts and "to manipulate information and ideas in ways that transform their meaning and implications" (Newmann & Wehlage, 1993, p. 9). Such tasks (e.g., opinion essays) exhibit a level of ambiguity because they do not have single, clear right or wrong answers (Doyle, 1983). Tasks that exhibit low levels of cognitive demand require students to reproduce knowledge from texts only by constructing literal summaries of events in texts or answering known-answer questions.

Students' opportunity to generate an extended or elaborated response in their written work by appropriately supporting their assertions with text-based evidence contributes to our conception of cognitive demand. As noted by Newell (2006), generating an extended response in writing is more difficult than filling out a worksheet or providing brief answers to questions because it demands more effort and thinking from the writer. Including this dimension in our consideration of cognitive demand allows for distinctions among cognitive operations as described by Doyle (1983). Opinion tasks and analyses can be high level and higher risk if

such a task holds students accountable for fully explaining and supporting their thinking. Conversely, such tasks can be low level and lower risk if students are not guided to explain their reasoning (e.g., provide an unsupported opinion) or use a single instance from a text in an analysis (e.g., provide an example of how a stage direction furthers an understanding of a character or action of a play). Low or nonexistent standards for evidence and constraints on student responses (e.g., providing a small box for students' writing) significantly lower a task's cognitive demand.

Finally, we consider in our conceptualization of task quality the texts to which students respond. Texts are central resources because they provide springboards for writing assignments. In their research on discussions that build students' reading comprehension skills, Beck and McKeown (2001) advocated for texts that are "conceptually challenging enough to require grappling with ideas and taking an active stance toward constructing meaning" (p. 10). Such texts should "provide the grist for children to explore ideas and to use language to explain ideas" (p. 14). The texts used as a basis for writing tasks need to contain sufficient grist to support interpretation and extended responses. Texts that are overly straightforward provide few opportunities for students to infer meaning, consider a perspective, or otherwise read between the lines of the written page.

Grist can be exhibited in multiple ways: in the complexity of subject matter content, plotlines, and themes. Grist can be exhibited in moral dilemmas (e.g., should people eat meat?) and real-life problems that do not have easy solutions (e.g., contending with the aftermath of hurricanes, lack of health care and clean water in some parts of the world). Texts that describe characters and historical figures that are neither completely good nor wholly bad also could be seen as exhibiting grist. Texts, or a collection of texts, need to contain enough ambiguity that they provide students with problems to grapple with in their writing, while supporting multiple perspectives.

We note that our conception of grist differs from the measures of text complexity included in the Common Core (Fisher, Frey, & Lapp, 2012; NGA Center & CCSSO, 2010). The Standards' model of text complexity has three components. The Lexile Framework for Reading provides a quantitative measure of text complexity based primarily on sentence length and word frequency. The second part of the model comprises a set of qualitative dimensions that include levels of meaning in text, text structures, knowledge demands, and the conventionality and clarity of language (Fisher et al., 2012; NGA Center & CCSSO, 2010). Our conception of grist differs from these measures because we are concerned with the features of texts that support the instructional goal of analytic writing, not those that directly impact students' comprehension.

Our measure is most aligned with the third component of the Common Core's model, reader and task considerations, which focuses on the use of text for particular instructional purposes. This component of the model, however, is considerably less specified than the first two (Hiebert & Mesmer, 2013). Teachers must use their professional judgment and subject matter knowledge to match texts to instructional tasks (and readers). The criteria that they are intended to use are not well defined, however, and provide little guidance to teachers deciding which texts to use in their curriculum for analytic writing assignments.

Research Questions

In the current study, we examine the quality of text-based writing tasks that teachers assign to their students and the contribution of task quality to students' analytic text-based writing. We focus on the tasks that teachers considered representative of their most challenging classroom work to provide insight on teachers' conceptions of high-level writing, as well as to establish a common basis for comparing teachers' tasks. We anticipated that teachers would more likely choose challenging tasks and call them typical. In our analyses, we hold constant other domains of literacy instruction that could increase the quality of students' responses, especially the amount of instruction focused on writing and reading comprehension. In addition to providing insight into students' text-based writing opportunities in classrooms, an understudied facet of writing research, the inclusion of our instructional covariates in our models also allows us to consider how different aspects of literacy instruction might differentially impact students' writing quality. The specific research questions that we address in our study are as follow:

1. What is the quality (cognitive demand and grist of text) of text-based writing tasks assigned to students?
2. Does the quality of assignment tasks, adjusting for other measures of literacy instruction, teacher characteristics, and student characteristics, predict overall classroom performance on a writing assessment and, in particular, different dimensions of students' text-based writing skills (analysis, evidence, organization, academic style, and mechanics)?

Methods

Sample

Fifth-grade classrooms ($n = 27$) from a longitudinal observational study of instructional quality form the

sample for this study. Although many teachers participated in the study for more than one year, we used only a year of classroom data per unique teacher to claim independence among classrooms in our statistical analyses.¹ For simplicity, when teachers had more than one year of data, we chose the first year the teacher participated in our study. All 27 classrooms were located in the same large urban district in Maryland. Seventy-eight percent of the teachers in our sample had a master's (61%) or doctoral (17%) degree. Additionally, 18% had advanced professional certification. The teachers' classroom experience ranged from two to 38 years, with an average of 17.

Information on student demographic characteristics, including prior year achievement on the state's accountability test, was collected from the district for all 793 students in our 27 classrooms. The demographic characteristics of students in our sample were roughly representative of the larger district. About 56% of the students received free (45%) or reduced-price (11%) lunch, and 11% had an individualized education plan. Students in our sample were predominantly minority, indicating the following group affiliations: African American (80%), Hispanic (12%), Native American (11%), Asian (5%), and white (3%).² In our prediction models, we adjusted for these student background characteristics in addition to teacher and classroom characteristics.

Measures and Procedures

Assignment Tasks

Over the course of an academic year, teachers were asked to provide six response-to-text assignments that they considered representative of challenging work. Teachers were provided with data collection materials at the beginning of the year. Assignments were collected in December ($n = 3$) and again in March ($n = 3$). Often, the most challenging tasks that teachers assign are culminating assignments given at the end of instructional units, so we gave teachers several weeks to assemble their tasks and student work.

Teachers described each assignment on a cover sheet and submitted the directions provided to students and four pieces of student work—two pieces of high quality and two pieces of medium quality—for each task. Teachers submitted a total of 149 tasks (92%, or an average of 5.52 tasks per teacher). Of these 149 tasks, we eliminated 11 that did not meet our sampling criteria because they were not text based. These tasks included creative writing assignments (e.g., asking students to write a two- to four-paragraph description that makes an everyday place seem terrifying) and personal narratives (e.g., asking students to write about a significant event in their lives). In addition, we were unable to locate the text for 15

assignments. Our final corpus of assignments comprises 123 tasks.

Assignments were rated using two scales from the Instructional Quality Assessment: cognitive demand and text grist (Junker et al., 2006; Matsumura et al., 2008).

- Cognitive demand is assessed on a 4-point scale (1 = low; 4 = high). Tasks that receive the lowest score of 1 guide students to recall isolated and fragmented facts about a text (i.e., the facts recalled about a text do not add up to a coherent representation of a text's gist or meaning) or write on a topic that is only loosely connected to a text. Tasks that receive a basic score of 2 guide students to construct a literal, surface-level representation of a text (e.g., recall the beginning, middle, and end of a text; compare characters on obvious differences, such as their appearance). Tasks that are rated a 3 guide students to interpret or analyze a text, but do not support students in fully developing their ideas or supporting their assertions with text-based evidence (e.g., write about the theme of a text in a few sentences, which allows little more than mere identification of the theme). Tasks that receive the highest score of 4 guide students to analyze or interpret a text (i.e., engage with the underlying meanings or nuances) and develop and support their assertions using evidence from the text.
- Grist of text used as the basis for the writing assignment is assessed on a 3-point scale (1 = low; 3 = high). Texts that receive the lowest score on this dimension are very short and straightforward and contain no or very little information for students to use in their writing. Texts that receive this score include short passages from assessments (e.g., test preparation materials), extremely short articles (one or two paragraphs) from children's magazines, highly patterned books that are designed for teaching print-sound code or fluency, or simple picture books intended for very young children.
Texts that receive a score of 2 contain some grist for meaning making but are limited in scope. These texts include excerpts from novels in the district's basal reader series that feature rich characters or events, but the potential of the text to support in-depth analytic writing was limited by the fact that only part of the novel was represented (e.g., development of a character was not shown). Texts that received a score of 2 also included short feature articles from children's magazines, short stories with straightforward plotlines, and picture books that contain some nuanced information.

Texts that receive the highest score for this dimension provide rich information for meaning making. These texts include many chapter books (e.g., *Where the Mountain Meets the Moon* by Grace Lin, *The Fear Place* by Phyllis Reynolds Naylor); short stories with nuanced content (e.g., “Eleven” by Sandra Cisneros); or informational texts that contain rich, newsworthy content and/or have content that supports multiple perspectives (e.g., a newspaper article describing a health care initiative for immigrants). Grist can also entail having students read multiple connected texts, such as short articles that provide multiple perspectives on a topic or short stories that share a common theme or are written by the same author.

Prior research examining the validity of the Instructional Quality Assessment assignment ratings indicates that task quality (a composite measure composed of grist, cognitive demand, and assessment criteria) predicted middle school students’ reading comprehension skills as assessed on the SAT-10 (Matsumura et al., 2008). The cognitive demand of assignments as assessed on the Instructional Quality Assessment also has been associated with variation in students’ use of academic language in their writing in the upper elementary grades (Crosson et al., 2012).

A graduate student researcher with prior experience as a high school English teacher rated all assignments and texts. To calculate inter-rater reliability, a subsample of 24 assignments (17% of the corpus) was assessed by a second rater for each dimension. Overall exact-scale

point agreement between raters was 92% and 79% for the cognitive demand and text grist scales, respectively. Cohen’s Kappa, which reports the exact match agreement adjusted for the distribution of scores across categories, was also calculated. Cohen’s Kappa of .84 and Pearson’s correlation ($r = .90$) indicate high agreement between raters overall on the cognitive demand scale. Cohen’s Kappa of .67 and Pearson’s correlation ($r = .80$) indicate moderately high agreement between raters on the grist scale.

Instructional Logs

Teachers self-reported the content and activities of their writing and reading comprehension instruction using online daily logs³ completed over a two- to three-week period in November, January, and May (30–45 days total). Participating teachers submitted over 85% of sampled days (mean [M] = 85.6% per teacher, standard deviation [SD] = 16.9%). Prior research shows that similar measures and data points form a reliable composite (Rowan, Camburn, & Correnti, 2004), are sensitive enough to detect differences among teachers’ instruction in schools implementing different comprehensive school reform programs (Correnti & Rowan, 2007), and are related to student learning (Correnti, Rowan, & Camburn, 2003). To prepare items for analysis, a data reduction technique was used to collapse individual log items into dichotomous indicators of whether a particular aspect of instruction was a focus on that day (see Table 1; for descriptions of component item groupings, see Correnti, 2007; Correnti & Rowan, 2007).

TABLE 1
Frequency (percentage of days strategy was covered) of Log Items Used to Construct Scales in Measurement Models

| Comprehension | Percentage | Writing | Percentage |
|---|------------|--|------------|
| Students discussing text | 51 | Direct teaching of writing skills | 35 |
| Checks for literal understanding | 50 | Prewriting activities | 25 |
| Direct teaching of comprehension skills and/or strategies | 48 | Teacher commenting on writing (i.e., conferring) | 25 |
| Students providing brief answers to questions | 45 | Practice writing | 19 |
| Students learning story structure | 43 | Integrating comprehension with writing | 16 |
| Students engaging in prereading activities | 40 | Writing multiple paragraphs | 15 |
| Students asked to analyze/evaluate text | 33 | Substantive revision | 13 |
| Focusing on writing and reading comprehension on the same day | 32 | Focusing on genre/literary techniques | 12 |
| Extending text comprehension activities through writing | 19 | Sharing writing with peers | 10 |
| Producing extended answers | 15 | Editing drafts | 9 |

Both a writing scale and a reading comprehension scale were generated as proxies to adjust for the amount of instructional time on each topic within literacy. Items that compose the writing scale include the amount of time teachers directly teach writing skills, engage students in process-oriented approaches (e.g., prewriting activities, editing and revising drafts, conferring about writing), and require students to generate multiparagraph responses. Items that compose the comprehension scale include the amount of time teachers engage students in text discussions, directly teach comprehension strategies, check for students' literal understanding of texts, teach story structure, and have students answer questions about what they read (for further detail on the log measures, see Correnti et al., 2012).

Computation of these scales was conducted using a (logistic) multilevel multivariate measurement model using HLM 7.1 (Raudenbush & Bryk, 2002). Dichotomous instructional items for a given day were nested in days, and days were nested in teachers. These models are useful for understanding psychometric properties of the scales (see, e.g., Raudenbush, Rowan, & Kang, 1991), adjusting for covariates in the model (e.g., time of year, item characteristics), and computing an empirical Bayesian residual to be used as a continuous indicator of the frequency of instructional practices. The measurement model indicated that significant variation existed among teachers on both the writing scale ($\chi^2 = 685.14$; $df = 111$) and the comprehension scale ($\chi^2 = 1,170.06$; $df = 111$).⁴ Teacher-level HLM reliabilities were .74 and .84, respectively, for the two scales.

Response-to-Text Assessment (RTA)

The RTA assesses students' ability to express their reasoning about academic texts and use text evidence to support their assertions in their writing (Correnti et al., 2012; Correnti, Matsumura, Hamilton, & Wang, 2013). The RTA was administered in one 60-minute session by teachers during the last week of May of each year of the study. Teachers read the text aloud to students while they followed along with their own copies. Teachers asked standardized questions at predetermined points in the text to aid students' literal comprehension of the text. Drawing on Questioning the Author techniques (Beck & McKeown, 2006), these standardized questions include initial questions with expected student responses and standardized follow-up questions if students do not respond as expected. The goal is to help ensure that students have a literal understanding of the text so they can address the prompt, while minimizing reading fluency and comprehension as potential sources of measurement error in the assessment of students' writing skills.

For fifth-grade classrooms, the RTA is based on a feature article from *Time for Kids* ("A Brighter

Future" by Hannah Sachs) about a United Nations effort to eradicate poverty in a rural village in Kenya. We chose this text because it represents the kind of informational text required by the Common Core. The text contains information that, in our judgment, could support extended writing (i.e., was rated a 2 for grist on our scale), but was short enough that it could be read aloud and discussed within a 15-minute period (leaving 45 minutes for students to generate their responses). The prompt asks students, "Based on the article, did the author provide a convincing argument that 'winning the fight against poverty is achievable in our lifetime'? Explain why or why not with 3–4 examples from the text to support your answer." Aligned with the Common Core (grade 5 Reading Standards), the prompt is intended to guide students to evaluate the author's argument and to encourage students to integrate multiple instances of appropriate textual evidence to support their assertions (NGA Center & CCSSO, 2010). Students' essays are assessed on five dimensions, each of which is rated on a 4-point scale (1 = low; 4 = high). Below we briefly describe each dimension (for a description of the individual RTA scale points and anchor papers, see Appendix A, which is available as supporting information for the online version of this article):

- Analysis assesses students' ability to demonstrate a clear understanding of the text's purpose and to make valid and perceptive claims that constitute an insightful response to the prompt.
- Evidence captures the degree to which students select and use details, including direct quotations from the text to support their key idea.
- Organization assesses the degree to which students' responses exhibit a strong sense of beginning, middle, and end and demonstrate logical flow between sentences and ideas.
- Academic style focuses on the extent to which the student writing demonstrated varied sentence lengths and complex syntactical structures, multiple uses of Tier 2 vocabulary (e.g., *appreciate*, *fortunate*, *instructions*, *miserable*), and correct application of sophisticated connectives (e.g., *however*, *meanwhile*).
- Mechanics/usage/grammar/spelling assesses students' ability to adhere to grade-appropriate standard writing conventions.

The RTA was scored by university student raters who underwent extensive training in using rubrics and annotated benchmark samples. Reliability checks for each study year included a minimum of 20% of

randomly sampled essays that were rated by a member of the research team. The inter-rater reliability was then calculated by examining the intraclass correlation (ICC) among the raters at each study year. ICCs at the dimension level were as follows: mechanics (.77), academic style (.77), analysis (.78), organization (.80), and evidence (.82). These ICCs indicate high agreement among raters in general and for each individual dimension (M. Graham, Milanowski, & Miller, 2012).

In prior work using only year 1 data ($n = 18$ fourth- and fifth-grade classrooms), we measured students' opportunities to learn analytic writing to demonstrate the utility of the opportunity structure as a predictor of variance in the average RTA performance when all five dimensions were combined (Correnti et al., 2012, 2013). We report a similar analysis here with all of our fifth-grade teachers (Cronbach's α across all five dimensions was .82). However, to further understand associations between students' learning opportunities and performance on the RTA, we also explore different patterns of writing instruction in relation to classroom performance on particular RTA dimensions. We predict that students in classrooms with higher cognitive demand assignments are more likely to have higher classroom performance on the analysis dimension than the mechanics/usage/grammar/spelling dimension.

Although student performance was low overall ($M = 2.11$ across dimensions), evidence suggests that the student essays varied in their performance on the different dimensions of writing. First, preliminary analyses examining the ICC of RTA dimensions nested within students and students nested within classrooms demonstrated that about half of the variance was between dimensions within students. This variance is also reflected in the correlations between the dimension scores of the RTA. Table 2 displays these correlations at the student level (upper diagonal) and at the teacher

level (lower diagonal). Therefore, we examined exploratory analyses to test our theory that different types of instruction would be differentially related, in predictable ways, to our RTA dimensions representing different constructs for young students' writing.

Maryland School Assessment (MSA)

The MSA is the state standardized test administered to all students in grades 3–8. The reading portion consists of 33 multiple-choice questions on vocabulary, word study, and reading comprehension and four brief constructed responses. Students' writing is assessed on a 4-point scale (0–3) depending on the extent to which their response addresses the “demands of the question” and “uses text-relevant information to show understanding” (Maryland State Department of Education, 2014, Score 2 section, para. 1). The overall test score entails three subscales: general reading, literary reading, and information reading. Two of these subscales (literary reading and information reading) were included in the present study as measures of students' prior achievement. Each of these subscales consists of nine multiple-choice items and two brief constructed responses.

Analyses

We used descriptive statistics to address our first research question about the quality of text-based writing tasks assigned to students. To address our second research question about the influence of assignment quality on students' analytic writing, we first aggregated assignment scores on the cognitive demand and grist-of-text dimensions to the teacher level (average of up to six sampled assignments over the year). The distribution of both the cognitive demand scores and the average grist-of-text scores are roughly normal. Using a three-level hierarchical linear model using HLM 7.1, we examined both the average cognitive demand and the average grist-of-text scores of assignments as a predictor of performance on the RTA. Below we describe a series of models with a progression of covariates to understand the influence of assignment quality on student learning both with and without other instructional covariates.

We began by examining a fully unconditional model (see column 2 of Table 4) and then included student background information and teacher and classroom characteristics in the model (for a description of the statistical model with all instructional predictors, see equations 1.1–1.3 in Appendix B, which is available as supporting information for the online version of this article). Next, while adjusting for student background, teacher characteristics, and classroom characteristics, we explored the influence of different measures of literacy instruction on classroom performance on the

TABLE 2
Bivariate Correlations Between Dimensions of Response-to-Text Assessment

| Dimension | 1 | 2 | 3 | 4 | 5 | 6 |
|--|-----|-----|-----|-----|-----|-----|
| 1. Analysis | — | .45 | .56 | .50 | .47 | .75 |
| 2. Evidence | .58 | — | .58 | .37 | .38 | .73 |
| 3. Organization | .76 | .74 | — | .58 | .52 | .83 |
| 4. Academic style | .71 | .48 | .90 | — | .64 | .78 |
| 5. Mechanics/usage/grammar/spelling | .71 | .58 | .80 | .86 | — | .80 |
| 6. Response-to-text assessment (average) | .85 | .79 | .95 | .89 | .90 | — |

Note. The upper diagonal represents student-level correlations, and the lower diagonal represents classroom-level correlations.

RTA. We examined comprehension instruction by itself (a single scale obtained from the multilevel multivariate measurement model of the log data described earlier (see column 4 of Table 4), and then we examined all of our instructional measures simultaneously (see column 5 of Table 4). Our primary interest in these models was the effect of assignment cognitive demand (β_{1013}) and grist of text (β_{1012}) on overall class performance on the RTA.

To further describe the empirical association between the assignment cognitive demand and grist-of-text scores, and class performance on the RTA, we analyzed component dimensions of the RTA to understand whether our writing measures had similar patterns of association across all five dimensions (for a description of the statistical model, see equations 2.1–2.3 in Appendix B, which is available as supporting information for the online version of this article). For each of the five dimensions of the RTA, we simultaneously examined all covariates, including all instructional variables and adjusting for all student and classroom characteristics. Our primary interest in these models was the effect of assignment cognitive demand (β_{1013}) and grist of text (β_{1012}) on class performance on the RTA for each dimension and whether the pattern of effects was similar to the pattern of effects for the writing scale from the log (β_{1011}). Comparing and contrasting associations between these different measures and patterns of associations on the five different dimensions can provide further insight into how different aspects of writing instruction relate to different aspects of students' writing performance.

Results

In this section, the results are organized by research question. First, we looked at the quality of the tasks (grist and cognitive demand) assigned to students and provide in-depth examples of different assignment types. To address our second question, we examined the relationship between the grist and cognitive demand of assignments and class performance on the RTA, including an analysis of different dimensions of academic writing (analysis, evidence, organization, academic style, and mechanics/usage/grammar/spelling), holding constant other aspects of literacy instruction.

Research Question 1

Quality of Text-Based Writing Tasks

Our results for our first research question indicate that the quality of teachers' text-based assignment tasks was low overall. Although the majority of texts on which the tasks were based (84.4%) exhibited some grist to

support an extended, analytical response (i.e., were rated a 2 or 3 for grist),⁵ only a very small percentage of tasks (23.8%) guided students to demonstrate complex thinking skills in their writing (i.e., received a score of 3 or 4 for cognitive demand). In other words, even when students were provided with texts that contained content that could support a high-level task, the large majority of assignments (76.2%) guided students to retrieve isolated facts about a text or engage with only surface-level features of texts in their writing (i.e., were scored a 1 [19.6%] or a 2 [56.6%], respectively, on this dimension).

Our results also showed that students had very few opportunities to produce extended writing. Only 10% of the tasks in our sample guided students to generate multiparagraph responses. The large majority of assignments guided students to produce very short responses (one or two sentences or sentence fragments; 23%) or generate a single paragraph of writing (67%). In the following subsections, we describe in more depth the tasks and student work that were representative of our corpus within each level of cognitive demand to provide further insight into students' text-based writing opportunities. Table 3 describes the frequency of assignments at each level of cognitive demand and, within these levels, the grist of text to which students responded.

Low Cognitive Demand: Recall of Isolated Facts and Writing on a Topic Loosely Connected to a Text

As shown in Table 3, about a fifth (19.7%, $n = 24$) of our assignment corpus received the lowest score for cognitive demand (rated a 1). The large majority of these assignments ($n = 21$) comprised worksheets or graphic organizers (e.g., Venn diagrams) that guided students to recall isolated and fragmented facts about a text. One worksheet, for example, guided students to answer a series of questions in response to the short story "La Bamba" by Gary Soto from the school's basal reader series. This story (rated a 2 for grist) is about a boy from a large family who, motivated to gain recognition among his siblings, enters his school talent show to lip-synch Ritchie Valens's song "La Bamba." In the middle of the boy's performance, the needle sticks on the record, leading to an unintentionally funny (and ultimately successful) conclusion. Five of the seven questions on the worksheet concerned events in the story (e.g., "What causes Manuel to enter the talent show?"), and two focused on vocabulary. The worksheet contained a short-answer question requiring a summary of a part of the story: "Describe Manuel's performance in the talent show." A response for this last question that was considered by the teacher as representative of high-quality work is "Manuel's

TABLE 3
Quality of Text-Based Assignment Tasks (N = 27 teachers, 123 assignment tasks)

| Cognitive demand rating | Grist of text rating | N | General descriptor | Sample assignment |
|-------------------------|----------------------|----|---|--|
| 1 (n = 24) | 1 | 5 | <i>Recall of isolated facts about low-level texts:</i> These tasks guide students to recall fragmented and isolated facts from simplistic texts (e.g., short comprehension test passages, one- or two-paragraph articles from children’s magazines, picture books intended for very young children). Or, the task guides students to write on a topic that is loosely connected to a simplistic text. | Students read a short text about Chinese and U.S. New Year celebrations in their basal reader and completed a graphic organizer comparing the two traditions. |
| | 2 | 11 | <i>Recall of isolated facts about mid-level texts:</i> These tasks guide students to recall fragmented and isolated facts about texts that contain some grist for meaning making (e.g., short stories or novel excerpts with straightforward content, many informational text selections from basal readers, feature articles from children’s magazines on noncontroversial topics). Or, the task guides students to write on a topic that is loosely connected to a text that contains some grist. | Students completed a worksheet in response to the short story “La Bamba” by Gary Soto. Five of the seven questions on the worksheet are multiple choice, and two require short answers (one or two sentences) recalling a single event in the story (e.g., “What happened when Manuel performed at the talent show?”). |
| | 3 | 8 | <i>Recall of isolated facts about high-level texts:</i> These tasks guide students to recall fragmented and isolated facts about texts that contain high levels of grist (e.g., Newbery Medal-winning chapter books, short stories with nuanced content, articles on newsworthy or controversial topics, multiple connected texts). Or, the task guides students to write on a topic that is loosely connected to a text that contains a high level of grist. | Students wrote one- or two-sentence descriptions of their favorite part of the novel <i>The Phantom Tollbooth</i> by Norton Juster. |
| 2 (n = 69) | 1 | 10 | <i>Surface-level engagement with low-level texts:</i> These assignments guide students to construct surface-level responses to simplistic text (e.g., short comprehension test passages, one- or two-paragraph articles from children’s magazines, picture books intended for very young children). | Students summarized a very short article (one or two paragraphs) from a children’s magazine. |
| | 2 | 38 | <i>Surface-level engagement with midlevel texts:</i> These assignments guide students to construct surface-level responses to texts that contain some grist for meaning making (e.g., short stories or novel excerpts with straightforward content, many informational text selections from basal readers, feature articles from children’s magazines on noncontroversial topics). | Students summarized the short story “La Bamba” by Gary Soto by identifying the main characters, setting, problem, and solution to the problem in a one-paragraph response. |
| | 3 | 21 | <i>Surface-level engagement with high-level texts:</i> These assignments guide students to construct surface-level responses to texts that contain high levels of grist (e.g., Newbery Medal-winning chapter books, short stories with nuanced content, articles on newsworthy or controversial topics, multiple connected texts). | Students described whether they like the story “Eleven” by Sandra Cisneros and why. |

(continued)

TABLE 3
Quality of Text-Based Assignment Tasks (N = 27 teachers, 123 assignment tasks) (continued)

| Cognitive demand rating | Grist of text rating | N | General descriptor | Sample assignment |
|-------------------------|----------------------|----|---|---|
| 3 (n = 23) | 1 | 4 | <i>Analysis and interpretation of low-level text (constrained response):</i> These assignments guide students to apply some level of analytic thinking to simplistic texts (e.g., short comprehension test passages, one- or two-paragraph articles from children’s magazines, picture books intended for very young children), but do not support students to elaborate on their thinking or provide evidence for their assertions. | Students identified the organizational pattern (e.g., description, problem/solution, compare/contrast) of a very short article from a children’s magazine (<i>Time for Kids</i>) and explained which alternate organization would work well for informing readers about the topic. |
| | 2 | 14 | <i>Analysis and interpretation of midlevel texts (constrained response):</i> These assignments guide students to apply some level of analytic thinking to texts that contain some grist for meaning making (e.g., short stories or novel excerpts with straightforward content, many informational text selections from basal readers, feature articles from children’s magazines on noncontroversial topics), but do not support students to elaborate on their thinking or provide evidence for their assertions. | Students read the poem “Secrets” by Myra Cohn Livingstone and wrote a short response to the question, “What big idea can you infer from reading the poem? Include specific words or phrases from the poem to explain your thinking.” |
| | 3 | 5 | <i>Analysis and interpretation of high-level texts (constrained response):</i> These assignments guide students to apply some level of analytic thinking to texts that contain high levels of grist (e.g., Newbery Medal-winning chapter books, short stories with nuanced content, articles on newsworthy or controversial topics, multiple connected texts), but do not support students to elaborate on their thinking or provide evidence for their assertions. | Students compared the main characters’ acts of courage in the short story “La Bamba” by Gary Soto and an excerpt from <i>The Fear Place</i> by Phyllis Reynolds Naylor. Students were guided to describe how the acts were similar (one paragraph) and different (one paragraph). |
| 4 (n = 6) | 1 | 0 | <i>Analysis and interpretation of low-level texts (elaborated response):</i> These assignments guide students to apply analytic thinking to simplistic texts (e.g., short comprehension test passages, one- or two-paragraph articles from children’s magazines, picture books intended for very young children), and support students to elaborate on their thinking and provide evidence for their assertions. | No examples exist from our corpus. |
| | 2 | 4 | <i>Analysis and interpretation of mid- to high-level texts (elaborated response):</i> These assignments guide students to apply analytic thinking to texts that contain at least some grist for meaning making (e.g., short stories or novel excerpts with straightforward content, many informational text selections from basal readers, feature articles from children’s magazines on noncontroversial topics), and support students to elaborate on their thinking and provide evidence for their assertions. | Students read the article “Are You a Screen Addict?” by Kristin Lewis from Scholastic’s <i>Scope</i> magazine and wrote a multiparagraph essay in which they convinced the reader to support or not support Screen-Free Week. |
| | 3 | 2 | <i>Analysis and interpretation of high-level texts (elaborated response):</i> These assignments guide students to apply analytic thinking to texts that contain high levels of grist (e.g., Newbery Medal-winning chapter books, short stories with nuanced content, articles on newsworthy or controversial topics, multiple connected texts), and support students to elaborate on their thinking and provide evidence for their assertions. | Students read multiple texts about the three major Western trails for pioneers and wrote a multiparagraph essay describing which of the trails they would take (Oregon, Mormon, or Santa Fe) if they were a settler moving to the West in the 1840s. Students were guided to give at least three reasons with detailed explanations for their choice. |

performance was good until the record player got stuck and kept repeating the same line so Manuel pantomimed the line like the record until Mr. Roybal ripped the needle across the record.”

This task received a rating of 1 because the questions posed to students were closed ended and only required recall of fragmented information from the reading. As shown in the example of student work that received a high grade for the assignment, the student could accurately retell Manuel’s performance in the talent show without grasping the significance of this scene in the text as a whole. The task would have received a higher rating if students had been guided to formulate a coherent representation of the text (e.g., to summarize the whole story) and/or consider the larger message of the story (e.g., the idea that seemingly negative events can have positive outcomes, or the irony of the story’s ending).

In contrast to assignments that guide students to recall fragmented, factual information from texts, a few assignments in our corpus ($n = 3$) received a rating of 1 for cognitive demand because they required students to write on a topic that was only loosely connected to the content of a text. One such assignment, for example, guided students to read an excerpt from the *The Fear Place* and write about a time that they faced a fear. This novel (rated a 3 for grist) focuses on the complex, competitive, and devoted relationship between two brothers. Here is an example of student work that the teacher considered to be of high quality:

At Six Flags me, my mom and my cousin were riding on the teacups. They made us dizzy and my mom said, “Let’s go eat.” We said “Yeah.” So after we ate we were face to face with my fear place “The Superman.” It is tall, big and fast. My fear place is a rollercoaster. My mom said “Let’s ride it.” My cousin said “yes.” I said “No! It is too big I will stay here and watch.” My mom said “let’s go.” I said scared “I will try it.” So I got on it and we rode it. We got off and I said “It’s not so bad.”

As illustrated in the student’s essays, students were not guided to engage with the content of the text, and indeed students need not have read the text to successfully complete the assignment. This assignment would have received a higher score (a 2) if students had engaged with the basic events in the texts and an even higher score (a 3 or 4) if it had guided students to engage with the rich content of the text in their responses (e.g., the difficult relationship between the brothers, the significance of the cougar to the story).

Basic-Level Cognitive Demand: Surface-Level Text Engagement

The majority of our assignment corpus (57%, $n = 69$) scored a 2 for cognitive demand. About a third ($n = 21$)

of these tasks required students to construct basic summaries. For example, one such assignment guided students to summarize the short story “La Bamba” by identifying the main characters, setting, problem, and solution to the problem. Here is an example of a student’s essay that the teacher considered to be a high-quality response:

In the story there is a kid named Manuel and he was the fourth of seven children. The main characters are Manuel, Ernie, Benny, Mr. Roybal. Manuel is excited to be in the talent show. But he just hopes that he doesn’t mess up. The next day Manuel goes to the talent show to perform when Manuel performs he sings but the record got jammed then it kept saying the same line over and over and Manuel kept singing the same line then everybody laughed. Then everyone said you were funny. Manuel [felt] very happy.

In contrast to the previously described task in response to the same story (rated a 1 for cognitive demand), this task received a higher score because it supported students in creating a coherent representation of the text that demonstrated understanding of cause–effect relationships in the text and the significance of a key event (i.e., Manuel’s performance in the talent show). This task did not receive a higher rating (a 3 or 4), however, because students were not supported in reaching beyond what was literally represented in the text. If, for example, students had been asked to infer the author’s message, this task would have received a higher score.

The remaining two thirds of the tasks that received a score of 2 for cognitive demand ($n = 48$) guided students to apply potentially more complex thinking processes in a superficial way. These tasks contained language in the directions that suggested that students would need to apply higher level thinking skills to complete the task successfully (e.g., compare and contrast characters or events, analyze settings and characters, describe how actions from a character changed the resulting action of a story, predict an ending, write opinions). Yet, the format of these assignments and expectations for completed student work guided students to address only obvious content in the text (e.g., compare characters on surface-level characteristics, such as their gender and hair color) and provide extremely limited explanation for their responses (e.g., write a single sentence predicting an ending of a story). The assignments provided little opportunity for students to engage deeply with the content of texts or construct new insights about what they read.

The following is an example of a task that we consider representative of a surface-level application of potentially more complex thinking skills. For this assignment, students read the story “Eleven.” This nuanced story (rated a 3 for grist) is about a girl who is

forced by her teacher to wear an ugly sweater. The story deals with complex ideas around getting older and how adults can humiliate and intimidate children. In the assignment directions, the teacher asked students to “evaluate [the] text, include textual evidence and support as well as personal thoughts and opinions.” Evaluation is a complex cognitive process involving judgment of the value or adequacy of something for a specified purpose. To make such judgments, the processes of comprehension and analysis must be undertaken. In this particular assignment, however, the purpose and the criteria for judgment were not specified. This is evident in the student work that was nominated by the teacher as representative of a high-quality response:

My opinion on the story Eleven by Sandra Cisneros about Rachel is that she is shy. She won't stand up for herself when Mrs. Price says it's her jacket when it's not. She is also a problem-solver. She thinks of ways to get rid of the jacket, like throwing it over the school fence. Rachel is also fragile because she breaks down and cries. She also feels embarrassed, because the jacket really isn't hers and the teacher makes her put it on.

Mrs. Price is pushy because she forces the jacket on Rachel, making her wear it. She's childish, because she doesn't even apologize after embarrassing Rachel. Mrs. Price is also mean, even though Rachel denies the jacket is hers, Mrs. Price tells Rachel it's her jacket.

I like how the author included much detail about how Rachel felt. I don't like how she didn't use more descriptive words to describe the teacher, and setting. I wish the author would've described how the kids in the class reacted to what was happening. I think overall the author did a good job with the problem and solution.

As shown in this essay (and other student work for this assignment), the majority of the student's response focused on summarizing the plot and describing the characters rather than evaluating the quality of the story. Although in the final paragraph the student began to judge elements of the writer's craft (e.g., the author's use of descriptive language), students were not guided to explain why more descriptive language for the teacher or setting would have improved the story or what was good about how the author handled the story's problem and solution. As a result, the task was equivalent to asking students to write about what they liked and did not like about the story, which is more reader response than analytical in nature.

Analyzing and Interpreting Text: Constrained Response

Only about a quarter of the tasks that we collected from teachers guided students to analyze or interpret a text, for example, by inferring themes from texts, supporting opinions with evidence, comparing themes and

characters, and analyzing story elements. The majority of these assignments received a score of 3 ($n = 23$), as opposed to a 4, because students' opportunity to produce an extended response was explicitly limited (e.g., students were provided with a small box in which to complete their answer or directed to limit their responses to one or two paragraphs). In one assignment that was rated a 3 for cognitive demand, students compared the acts of courage carried out by characters in *The Fear Place* and the short story “La Bamba.” Students completed a Venn diagram to help plan their essays and then wrote a two-paragraph response. Here is an example of a student's response that the teacher deemed high quality:

Both Manuel's and Doug's act of courage are alike because they both did something they were afraid of. Also their acts of courage are alike because they both thought of stopping, but they didn't stop, they went on. The last reason their acts of courage are like because they both became brave and they became less nervous and scared why they were doing the thing they feared.

Doug's act of courage is different from Manuel's because Doug had someone to help him. Also Manuel's act of courage is different from Doug's because Manuel wasn't scared during all of the show, well Doug was scared...all the time while he was doing the thing he feared. Another difference is Manuel's act of courage came from wanting to be in the limelight and Doug's came from wanting to make up with his brother. Another difference is Doug has to use a lot more effort because he knew what his fear felt like cause he had done it before, but Manuel didn't have to use much effort because his fear wasn't as big as Doug's.

In this task, students engaged in the cognitive process of comparing and contrasting. They were not simply asked to compare the two characters generally, on concrete and potentially surface qualities; rather, students were to consider the types of courage exhibited by the two characters. This focus elevates the task because students must engage in deeper analysis of the text to make inferences about the characters' actions and motivations. This task falls short of receiving the highest cognitive demand score, however, because it is unclear what purpose the comparison/contrast serves. The exercise of comparing two characters does not appear to lead to a larger understanding of the meaning of the texts.

Analyzing and Interpreting Text: Elaborated Response

Only six assignments from our total corpus received the highest score for cognitive demand (a 4), and of these assignments, only two appear to meet the expectations for analytic text-based writing expressed in the Common Core for fifth-grade writing. One of these assignments guided students to read the article “Are You a Screen Addict” by Kristin Lewis from Scholastic's *Scope*

magazine (rated a 2 for grist) and write an essay in which they convinced the reader to support or not support Screen-Free Week. The second assignment guided students to read and synthesize multiple text selections (e.g., from their English language arts anthology, social studies textbook, and an Internet resource) about the three major Western trails for pioneers.

Considered individually, the majority of these texts would have received a rating of 2 for grist. As a set, however, the texts contained content that could support a substantive response (i.e., a rating of 3 for grist). The assignment task guided students to draw on the content of these texts in writing a multiparagraph essay in response to the following prompt: "Suppose you are a settler moving west in the 1840s. Which of the western trails would you take (Oregon Trail, Mormon Trail, or Santa Fe Trail) and why? Give at least three reasons with detailed explanations." The following is an example of a student essay that was considered by the teacher as representative of high-quality student work:

Have you ever heard of the Mormon Trail? If I were a pioneer in the 1840's, I would have picked the Mormon Trail as my favorite trail. I really wish I could have been there to travel the Mormon Trail. Here is why.

One reason I chose the Mormon Trail is because unlike the Santa Fe Trail and the Oregon Trail, the Mormon Trail is only 1,300 miles long, not 2,000 miles like the Oregon Trail or the Santa Fe trail that took eight weeks to travel.

Another reason that I would choose the Mormon Trail is that people who were traveling on that trails were going west for religious freedom. That supports my belief that all people are different and they shouldn't have to believe what somebody else believes in. If I wanted religious freedom, I would have traveled with the people on the Mormon Trail.

The last reason that I would take the Mormon Trail is that the trail started in Illinois, went across the Great Plains and over the Rocky Mountains. I would like to go on that trail because I like to explore different places. Traveling the Mormon Trail in this area was not as bad as the other trails that had harsh heat, storms, and fights with the Indians.

I think that the Mormon rail would be my best choice to travel on because you don't have lots of the dangers that the other trails had. You don't have to worry about too much danger or a long and difficult journey.

In contrast to less cognitively demanding compare-and-contrast tasks (e.g., complete a Venn diagram, compare characters on obvious differences), this task received a high rating because it supported students in using an evaluative frame to consider the merits of the various trails and synthesize information to form original conclusions. The comparisons that students were guided to make led to a judgment and decision rather than serving as ends in themselves. This task could be seen as both authentic and rigorous intellectual work.

Research Question 2

Effect of Assignment Cognitive Demand on RTA Performance

Our second research question focused on whether the quality of assignment tasks (grist and cognitive demand) is associated with students' analytic text-based writing skills beyond other measures of instruction. Our results indicate that the cognitive demand of teachers' assignments predicted multiple features of students' writing performance, including their ability to reason analytically about texts, use evidence to support their claims, and organize their writing. Grist of text, in contrast, significantly predicted only one feature of student writing, use of evidence, and was marginally significant for the organization of students' essays.

Effect of Assignment Task Quality on Classroom RTA Performance

Table 4 describes the findings for our progression of analytic models examining instructional effects on class RTA performance (an average score across all five dimensions). The second column provides the ICC for the fully unconditional three-level model with dimensions nested in students and students nested in classrooms. Notably, about half of the variance lies within students between dimensions. Thus, students did not score uniformly across dimensions (a point that we have discussed elsewhere; see, e.g., Correnti et al., 2013). Of the remaining variance, approximately 25% lies between classrooms, whereas the remainder is between students within classrooms.

Column 3 presents findings from a model with student and classroom background information. Examining the random-effects portion of this model, the addition of these covariates (prior to instructional covariates being added to the model) explains over half of the variance between classrooms and about a third of the variance between students within classrooms. Column 3 shows that in classrooms taught by teachers with a PhD, classroom performance on the RTA was about 0.38 points higher, adjusting for all other background factors. Column 3 also includes the extent to which teachers' culminating assignments were based on fiction texts as opposed to nonfiction. Recall that the fifth-grade RTA used in this study was based on a nonfiction text, and we wanted to adjust for any practice effects on performance due to prior exposure to nonfiction texts and prompts. The coefficient for the percentage of fiction texts is significant and positive. In other words, in classrooms with less exposure to nonfiction texts, on average, there was an associated higher classroom performance on the RTA. This effect subsequently disappears after we account for different measures of writing

TABLE 4
Effects of Instructional Variables on Classroom Average Performance on the Response-to-Text Assessment

| Classroom- and student-level variables | Fully unconditional model | | Background only | | Comprehension only | | All ^a | |
|---|---------------------------|------|-----------------|------|--------------------|------|------------------|------|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| <i>Classroom level (n = 27)</i> | | | | | | | | |
| Intercept, β_{100} | 2.09*** | 0.06 | 2.18*** | 0.12 | 2.18*** | 0.12 | 2.17*** | 0.11 |
| Master's degree, β_{101} | | | -0.13 | 0.14 | -0.12 | 0.14 | -0.12 | 0.11 |
| PhD, β_{102} | | | 0.38** | 0.11 | 0.38** | 0.12 | 0.40* | 0.17 |
| Advanced professional certification, β_{103} | | | 0.04 | 0.12 | 0.04 | 0.12 | 0.09 | 0.10 |
| Curricular coordinator, β_{104} | | | 0.07 | 0.15 | 0.06 | 0.15 | 0.01 | 0.13 |
| Class percentage receiving free lunch, ^b β_{105} | | | 0.01 | 0.04 | 0.01 | 0.04 | -0.02 | 0.05 |
| Class percentage receiving reduced-price lunch, ^b β_{106} | | | -0.05 | 0.04 | -0.06 | 0.04 | -0.05 | 0.05 |
| Years of experience, ^b β_{107} | | | 0.06 | 0.05 | 0.04 | 0.05 | 0.05 | 0.04 |
| Number of English language arts methods courses, ^b β_{108} | | | -0.03 | 0.04 | -0.04 | 0.04 | -0.05 | 0.04 |
| Percentage of fiction texts, ^b β_{109} | | | 0.11* | 0.04 | 0.11* | 0.04 | 0.05 | 0.05 |
| Comprehension factor, ^b β_{1010} | | | | | 0.03 | 0.05 | -0.15† | 0.08 |
| Writing factor, ^b β_{1011} | | | | | | | 0.21* | 0.08 |
| Average grit of text, ^b β_{1012} | | | | | | | 0.12 | 0.11 |
| Average cognitive demand, ^b β_{1013} | | | | | | | 0.25* | 0.08 |
| <i>Student level (n = 793)</i> | | | | | | | | |
| Male, β_{110} | | | -0.16*** | 0.03 | -0.16*** | 0.03 | -0.16*** | 0.04 |
| Hispanic, β_{120} | | | 0.10 | 0.07 | 0.10 | 0.07 | 0.09 | 0.08 |
| Native American, β_{130} | | | -0.10 | 0.10 | -0.10 | 0.10 | -0.09 | 0.08 |
| Asian, β_{140} | | | 0.32*** | 0.10 | 0.32*** | 0.10 | 0.30** | 0.11 |
| Black, β_{150} | | | 0.10 | 0.09 | 0.10 | 0.08 | 0.11 | 0.07 |
| Receiving free lunch, β_{160} | | | -0.13** | 0.05 | -0.13** | 0.05 | -0.13** | 0.04 |
| Receiving reduced-price lunch, β_{170} | | | -0.08 | 0.05 | -0.08 | 0.05 | -0.08 | 0.06 |
| Age in years, ^b β_{180} | | | 0.00 | 0.04 | 0.00 | 0.04 | 0.01 | 0.04 |
| Absences, ^b β_{190} | | | -0.07*** | 0.02 | -0.07*** | 0.02 | -0.06*** | 0.02 |
| Prior achievement on the literary scale, ^b β_{1100} | | | 0.14*** | 0.03 | 0.14*** | 0.03 | 0.14*** | 0.02 |
| Prior achievement on the informational scale, ^b β_{1110} | | | 0.13*** | 0.03 | 0.13*** | 0.03 | 0.13*** | 0.02 |
| <i>Random effects</i> | | | | | | | | |
| Between-classroom variance (τ_{π}) | 0.082 | | 0.039 | | 0.037 | | 0.020 | |
| Between-student variance (σ_e^2) | 0.248 | | 0.159 | | 0.159 | | 0.159 | |
| Between-dimension variance (σ_e^2) | 0.335 | | 0.335 | | 0.335 | | 0.335 | |

Note. SE = standard error.

^aThe mean across all dimensions for students taking the response-to-text assessment was $\mu = 2.12$ (standard deviation = 0.64). The raw standard deviation across students was used as the denominator to calculate effect sizes.

^bContinuous variables standardized and centered around their grand mean.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

TABLE 5
Effects of Instruction Variables on Dimension Scores on the Response-to-Text Assessment

| Classroom- and student-level variables | Analysis ^a | | Evidence ^a | | Organization ^a | | Academic style ^a | | Mechanics/usage/grammar/spelling ^a | |
|--|-----------------------|------|-----------------------|------|---------------------------|------|-----------------------------|------|---|------|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| <i>Classroom level (n = 27)</i> | | | | | | | | | | |
| Intercept, β_{p00} | 1.74*** | 0.13 | 2.10*** | 0.16 | 2.18*** | 0.16 | 2.29*** | 0.13 | 2.49*** | 0.19 |
| Master's degree, β_{p01} | 0.02 | 0.11 | -0.11 | 0.14 | -0.16 | 0.16 | -0.13 | 0.11 | -0.20 | 0.17 |
| PhD, β_{p02} | 0.29 | 0.18 | 0.78** | 0.22 | 0.57* | 0.24 | 0.26 | 0.18 | 0.10 | 0.27 |
| Advanced professional certification, β_{p03} | 0.04 | 0.10 | 0.00 | 0.12 | 0.00 | 0.14 | 0.02 | 0.10 | 0.38* | 0.15 |
| Curricular coordinator, β_{p04} | -0.02 | 0.13 | 0.07 | 0.16 | -0.06 | 0.18 | -0.07 | 0.13 | 0.13 | 0.20 |
| Class percentage receiving free lunch, β_{p05} | -0.05 | 0.05 | 0.00 | 0.06 | 0.02 | 0.06 | 0.00 | 0.05 | -0.08 | 0.07 |
| Class percentage receiving reduced-price lunch, β_{p06} | -0.09 | 0.05 | -0.07 | 0.06 | -0.02 | 0.07 | -0.01 | 0.05 | -0.05 | 0.08 |
| Years of experience, β_{p07} | 0.00 | 0.04 | 0.00 | 0.05 | 0.12 [†] | 0.06 | 0.09 [†] | 0.04 | 0.02 | 0.07 |
| Number of English language arts methods courses, β_{p08} | -0.02 | 0.04 | -0.09 | 0.05 | -0.03 | 0.06 | 0.00 | 0.04 | -0.10 | 0.07 |
| Percentage of nonfiction texts, β_{p09} | 0.06 | 0.05 | 0.10 | 0.06 | 0.05 | 0.07 | 0.03 | 0.05 | 0.01 | 0.07 |
| Comprehension factor, β_{p010} | -0.01 | 0.08 | -0.30** | 0.09 | -0.27* | 0.11 | -0.10 | 0.08 | -0.10 | 0.12 |
| Writing factor, β_{p011} | 0.06 | 0.08 | 0.27* | 0.10 | 0.31* | 0.11 | 0.15 [†] | 0.08 | 0.24 [†] | 0.12 |
| Average grist of text, β_{p012} | -0.05 | 0.11 | 0.33* | 0.13 | 0.28 [†] | 0.15 | 0.15 | 0.11 | -0.11 | 0.16 |
| Average cognitive demand, β_{p013} | 0.31** | 0.09 | 0.41** | 0.10 | 0.28* | 0.12 | 0.01 | 0.09 | 0.27 [†] | 0.13 |
| <i>Student level (n = 793)</i> | | | | | | | | | | |
| Male, β_{p10} | -0.11* | 0.05 | -0.17* | 0.07 | -0.22*** | 0.06 | -0.11* | 0.05 | -0.21** | 0.07 |
| Hispanic, β_{p20} | 0.12 | 0.11 | 0.12 | 0.13 | 0.11 | 0.12 | 0.03 | 0.11 | 0.08 | 0.14 |
| Native American, β_{p30} | -0.09 | 0.11 | -0.02 | 0.13 | -0.07 | 0.12 | -0.11 | 0.12 | -0.16 | 0.15 |
| Asian, β_{p40} | 0.10 | 0.15 | 0.33 [†] | 0.17 | 0.37* | 0.16 | 0.28 [†] | 0.15 | 0.45* | 0.19 |
| Black, β_{p50} | 0.13 | 0.10 | 0.24* | 0.12 | 0.15 | 0.11 | -0.06 | 0.10 | 0.13 | 0.13 |
| Receiving free lunch, β_{p60} | -0.10 [†] | 0.06 | -0.13* | 0.07 | -0.07 | 0.06 | -0.13* | 0.06 | -0.20** | 0.07 |
| Receiving reduced-price lunch, β_{p70} | -0.14 | 0.09 | 0.03 | 0.10 | -0.03 | 0.09 | -0.16 [†] | 0.09 | -0.08 | 0.11 |
| Age in years, β_{p80} | 0.02 | 0.05 | 0.08 | 0.06 | -0.03 | 0.06 | -0.01 | 0.05 | 0.01 | 0.07 |
| Absences, β_{p90} | -0.06* | 0.03 | -0.06 [†] | 0.04 | -0.08* | 0.03 | -0.04 | 0.03 | -0.08* | 0.04 |
| Prior achievement on the literary scale, β_{p100} | 0.14*** | 0.03 | 0.15*** | 0.04 | 0.10** | 0.03 | 0.11*** | 0.03 | 0.19*** | 0.04 |
| Prior achievement on the informational scale, β_{p110} | 0.07* | 0.03 | 0.10** | 0.04 | 0.16*** | 0.03 | 0.12*** | 0.03 | 0.19*** | 0.04 |

(continued)

TABLE 5
Effects of Instruction Variables on Dimension Scores on the Response-to-Text Assessment (continued)

| Classroom- and student-level variables | Analysis ^a | | Evidence ^a | | Organization ^a | | Academic style ^a | | Mechanics/usage/grammar/spelling ^a | |
|---|-----------------------|----|-----------------------|----|---------------------------|----|-----------------------------|----|---|----|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| <i>Random effects</i> | | | | | | | | | | |
| Fully unconditional model | | | | | | | | | | |
| Between-classroom variance (τ_{π}) | 0.067 | | 0.118 | | 0.139 | | 0.059 | | 0.134 | |
| Between-student variance (σ_e^2) | 0.145 | | 0.349 | | 0.245 | | 0.170 | | 0.569 | |
| Variance (percentage explained) | | | | | | | | | | |
| Conditional model | | | | | | | | | | |
| Between-classroom variance (τ_{π}) | 0.011 (81%) | | 0.019 (83%) | | 0.038 (73%) | | 0.013 (80%) | | 0.040 (70%) | |
| Between-student variance (σ_e^2) | 0.096 (33%) | | 0.260 (26%) | | 0.153 (37%) | | 0.104 (38%) | | 0.387 (32%) | |

Note. SE = standard error.

^aRaw means and standard deviations (SDs) across students for each dimension were as follows: analyze ($\mu = 1.79$; $SD = 0.68$), evidence ($\mu = 2.15$; $SD = 0.89$), organization ($\mu = 2.09$; $SD = 0.79$), style ($\mu = 2.05$; $SD = 0.66$), and mechanics/usage/grammar/spelling ($\mu = 2.40$; $SD = 1.01$). The raw SD was used to calculate effect sizes.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

instruction.⁶ No other teacher or classroom covariates predict classroom performance on the RTA.

Column 4 adds one covariate from the log measuring teachers' comprehension instruction. The coefficient for the comprehension scale is positive but not significant. There is approximately a 4% reduction in variance between classrooms as a result of adding this nonsignificant instructional covariate versus the model with background characteristics alone.

Findings for the effect of teachers' average assignment task quality on classroom average RTA performance are presented in column 5 of Table 5. Although the average gist of the text is nonsignificant, the average cognitive demand of the assignments is predictive of classroom RTA performance. After adjusting for student background, teacher characteristics, and aggregate classroom characteristics, the effect of assignment cognitive demand is significant when adjusting for other writing instruction covariates (column 5: $\beta_{1013} = 0.25$, effect size [ES] = 0.39). These findings show that in classrooms where the assignment quality was higher by a full point (e.g., from 1 to 2, from 1.5 to 2.5), classroom performance on the RTA was about four tenths of a standard deviation higher, holding everything else in the model constant.

It is noteworthy that the effects of the writing scale measure developed from the log data are also a significant predictor of classroom RTA performance (column 5: $\beta_{1011} = 0.21$, ES = 0.33).⁷ Thus, each of two writing

measures explains a unique portion of the variance in classroom RTA performance. Only one other classroom characteristic is a significant predictor of classroom RTA performance. Once measures of writing are adjusted for in the model, whether a teacher has obtained a PhD is a significant and positive predictor of classroom RTA performance. Variance components from column 5 show that more than three quarters of the original variance between classrooms is explained by this model.

Effect of Assignment Task Quality on RTA Dimension Scores

Table 5 displays the findings for all of our covariates on individual RTA dimension scores, estimated simultaneously (see the model in Appendix B, which is available as supporting information for the online version of this article). Although both assignment cognitive demand and the writing scale from the log predict a unique portion of the variance in classroom RTA performance, each measure has a different pattern of results for individual RTA dimension scores. For example, the effect of assignment cognitive demand is significant on the analysis dimension (column 2: $\beta_{1013} = 0.31$, ES = 0.46), whereas the effect of the writing scale from the log is not (column 2: $\beta_{1011} = 0.06$, n.s.). The effect of assignment cognitive demand is prominent for the analysis (ES = 0.46), evidence (ES = 0.46), and organization (ES = 0.35) dimensions. Effects for the writing scale are seen broadly on all of the

dimensions with the exception of analysis, including evidence (ES = 0.30), organization (ES = 0.39), and academic style (ES = 0.23). Thus, although both measures predict the evidence and organization dimensions, the assignment cognitive demand is uniquely predictive of classroom performance on the analysis dimension, and the writing scale from the log is uniquely predictive of academic style. Both are also marginally significant on mechanics (ES = 0.27 and 0.25, respectively, for assignment cognitive demand and the writing scale).

The grist of the text is our third writing covariate with a distinct pattern of association with our five different dimensions of the RTA. After adjusting for writing and reading comprehension instruction, the average grist of the text was associated with higher classroom performance on evidence (ES = 0.37) and a marginally significant association for higher classroom performance on organization (ES = 0.35). This pattern is logical, as the grist of the text only has influence on those aspects of writing directly affected by how students used evidence from the text to build coherent support for their ideas. Variance components from this model reveal findings similar to Table 4; that is, 26–40% of the between-student variance is explained by this model (varying slightly by dimension), and roughly three quarters of the between-classroom variation is explained by this model.

Discussion

Past research and policy recommendations have emphasized increasing the amount of time that students spend engaged in writing, and provided robust evidence supporting process-oriented forms of instruction (see studies summarized in S. Graham & Perin, 2007). Questions remain about the nature of instruction that develops different aspects of students' text-based writing skills. Holding constant other aspects of literacy instruction, our results provide evidence that the cognitive demand of text-based writing assignments predicts multiple features of students' writing performance, including their ability to reason analytically about texts in their writing. Aligned with past policy recommendations (NCW, 2003), the amount of instructional time focused on writing also independently predicted the quality of students' essays. Our results strongly suggest, however, that cognitively demanding tasks uniquely help develop students' higher level thinking about texts in their writing. Therefore, they should be considered an essential part of writing instruction reform to meet the ambitious goals of the Common Core.

We also were concerned with understanding the impact of the texts that students wrote about in classrooms (i.e., their level of grist) on their writing

outcomes, based on the assumption that the texts to which students respond set the conditions for intellectual work. Looking descriptively at our corpus, this assumption appears to be correct, in that most tasks that received high scores for cognitive demand drew on higher quality texts as springboards for students' thinking and writing. Our regression analyses indicate, however, that the grist-of-text rating differs from our other two writing measures. Its influence appears more localized and is not significant on the broader RTA measure.

After adjusting for the influence of other reading and writing instruction, the grist of the text is most likely to matter for one dimension of students' writing: their ability to extract important evidence from a text and discuss it with specificity and elaboration. Overall, however, the balance of evidence from our study suggests that how tasks position students to engage with the content of texts is more influential for students' writing skill development than the quality of texts on their own. Placing high-quality texts in classrooms likely will create a very limited (if any) impact on realizing the writing goals set out in the Common Core, especially if absent efforts to ensure students are provided with writing tasks that support high-level thinking about what they read.

Most importantly, our study draws attention to the very low level of cognitive demand in the text-based writing tasks assigned to students. This is not a new finding. Researchers have noted that students have few opportunities to produce extended responses and engage in intellectual problem solving in their writing (Applebee & Langer, 2011; Gilbert & Graham, 2010; S. Graham, Cappizi, Harris, Hebert, & Murphy, 2014). Our study is important, however, for pointing to particular limitations on how student writers are supported in engaging with text content. Regardless of the quality of the text to which students respond, our results show that students are almost always guided to recall basic facts or recapitulate surface-level events. Students rarely receive assignments that require them to analyze and interpret what they read, explain their thinking, and support their claims with textual evidence. In other words, opportunities for students to practice the analytic forms of writing that support text comprehension (S. Graham & Hebert, 2010) and knowledge of subject matter content (e.g., Monte-Sano & De La Paz, 2012) and put students on the path to college readiness and success are exceedingly rare.

Implications for PD

Efforts to close the gap between the Common Core's vision of text-based writing and the tasks that teachers most often assign to students will need to include a

significant investment in teachers' PD. In this section, we discuss some of the potential features of PD that would support transformation in teachers' writing assignments. As a starting point, professional developers likely will need to devote significant attention to building teachers' pedagogical knowledge by guiding them to study features of tasks that vary in cognitive demand (see, e.g., Matsumura, 2005, for a collection of writing tasks for various grade levels that have been rated and annotated). Following this exercise, teachers could collaboratively practice assessing sets of common tasks, commenting on what makes them cognitively challenging or how each task might be improved. In the next step, teachers could be guided to examine their own tasks, redesigning them as necessary to align with features that support cognitive rigor.

Because efforts to transform instruction often must help teachers significantly change their often tacit assumptions about teaching and learning (Borko, Mayfield, Marion, Flexer, & Cumbo, 1997), transforming practices around task design also would benefit from a consideration of teachers' existing beliefs about effective writing tasks and instruction generally. Professional developers could work with teachers to surface, and challenge, the connections that teachers make among task design, assignment directions, and terms found in the Common Core (Spillane, Reiser, & Reimer, 2002). Teachers may benefit from explicit unpacking and rethinking of words such as *analyze* and *evaluate* and may need to build a clearer, more rigorous understanding of what it means to have students support claims with evidence. Such intervention might help teachers clarify for themselves and communicate to students the expectations for successful completion of the task. Teachers then may learn to anticipate how students will approach the task in ways that are less cognitively demanding than intended (e.g., summarizing the text vs. analyzing its theme), and begin to proactively take measures to ensure students' high cognitive engagement with the task (e.g., more precise assignment directions or instruction related to the task).

PD interventions that support text-based writing might also benefit from a component related to text selection (which includes both print and nonprint media). The Common Core's model of text complexity leaves it to teachers to choose texts that fit particular instructional goals (and readers' interests). This component of the Common Core's text complexity model could be improved by further elaboration. We believe that our conception of grist could serve this purpose by providing criteria that teachers and professional developers could use to evaluate the potential of texts to support analytic writing. Working together, teachers and professional developers could

consider, for example, features of texts that are not often contemplated in measures of text complexity, such as ambiguity of characters' motives, complex characters and relationships (e.g., characters that are both good and bad, sibling rivalry and devotion), true moral dilemmas, and problems in life that do not have clear-cut solutions. Such features of texts lend themselves to analytic writing because they present problems that writers can grapple with, and support multiple points of view (i.e., conditions where reasonable minds can disagree on a topic). This latter point is important for empowering students to develop their own opinions that might differ from those of their classmates and teachers (Yeh, 1998).

Limitations and Directions for Future Research

Our study has several limitations. First, the RTA has not been extensively studied with multiple populations of students. The inferences that we made about the quality of students' writing on the basis of the RTA scores may not apply to all groups of students (e.g., English learners). Second, although we accounted for prior achievement in our models (reading and, to some extent, writing as measured by brief constructed responses in the MSA), we lacked the resources to administer multiple pre- and posttest assessments of students' analytic text-based writing. Further research is needed that considers change over time in students' writing skills as an outcome measure with a larger and more diverse sample to strengthen causal inferences about the role of tasks and the development of students' writing competencies.

In interpreting our results, it is also important to bear in mind the limitations of the texts to which students responded. The nonfiction texts that teachers chose and that comprised our assignment corpus were nearly all short articles focused on high-interest topics (e.g., bedbugs, how bees communicate, biographical information about a famous person). These texts were not squarely in a discipline outside of English language arts. Future research should examine how assignments within disciplines such as science and history influence students' ability to respond to text in writing.

Research that provides further insight on the role that other domains of literacy instruction play in the development of students' text-based writing skills is needed. Students can respond to a cognitively demanding task in ways that are not indicative of deep analysis. How students engage with a task, therefore, depends on prior teaching, the specific instructions provided for a given task, and teachers' standards for what counts as a quality response. Because generating

high-quality text-based arguments requires students to have a deep understanding of what they read, the quality of classroom text discussion likely also plays an important role. Instruction that develops analytic text-based writing will need to focus on developing students' reading comprehension and argumentation skills (e.g., how to formulate arguments and articulate reasons explaining what counts as evidence; Newell, Beach, Smith, & VanDerHeide, 2011; O'Hallaron, 2014), in addition to teaching the skills and strategies of effective writers. More research is needed that provides insight on the combination of instructional behaviors that support students in writing successfully in response to cognitively demanding text-based writing tasks.

Our study raises many questions about what teachers believe is good- or acceptable-quality text-based writing. We asked for samples of student work that teachers believe are representative of high- and medium-quality writing for a given assignment to provide insight into the task design (i.e., constraints, affordances) and written directions given to students. Although the quality of student work for individual tasks was not an explicit focus of our investigation, we noted that student writing that was considered by teachers as representative of high-level work often did not show strong evidence of analytic thinking and reasoning. Further research must systematically look at the interrelationship of task design, teachers' expectations for high-quality student responses, and students' analytic text-based writing skill development.

Although our call for action on the basis of our study's findings focuses on teachers (e.g., the need for increased PD), we emphasize that factors beyond teachers' control also play a critical role in assignment quality. Teachers frequently feel pressured to align their writing instruction to their state's accountability test (Applebee & Langer, 2013; Koretz & Hamilton, 2006). Although some states have adopted assessments that are aligned to the Common Core (i.e., are part of the Partnership for Assessment of Readiness for College and Careers or the Smarter Balanced consortium), many continue with their previous state assessments that do not foreground analytic text-based writing. In these states, the vision of desirable writing tasks promoted in the Common Core likely will compete with the reality of test-based writing for which teachers are held accountable. Teachers can be required by their district to use a curriculum that does not necessarily support analytic text-based writing (see, e.g., Escher, 2015). Significantly improving students' analytic text-based writing opportunities will require changes beyond instructional improvements at the teacher level, into multiple other aspects of the education system.

NOTES

The research reported here was funded through grants from the William T. Grant Foundation and the Spencer Foundation. The opinions expressed in this article are those of the authors, not the sponsors. The authors remain responsible for any errors in the work.

¹ Inferences from our analyses do not change when we examine models including all 52 fifth-grade classrooms in the study corpus. These analyses sharpen the statistical findings because the effect size increases simultaneously with the power. We also have not included any classrooms with imputed assignment scores because we wanted strict correspondence between the descriptive analyses of the assignments themselves and the assignments used in our empirical prediction models.

² These percentages total 111% because the school district reported more than one ethnicity for some students (e.g., white and Asian).

³ The online version of the logs for this study included a scroll-over feature that provided the glossary definition for each item.

⁴ These measurement models were run using all 52 fifth-grade classrooms available across the three years of the study. These include the 27 classrooms that were a focus for this study (i.e., corresponding to the unique teachers who contributed the assignments described in Table 4).

⁵ Two thirds of the texts to which students responded in our corpus were fiction, and one third was nonfiction (informational). This proportion of fiction to nonfiction texts is aligned with other research analyzing the text materials available to students in the upper elementary grades (Braker-Walters, 2014).

⁶ In subsequent analyses, we observed that the average grist of text was significantly higher for fiction ($\mu = 2.25$) versus nonfiction ($\mu = 1.95$) texts ($t = 2.39$, $df = 117$, $p = .019$). The average cognitive demand of assignment tasks was not significantly different ($t = 1.01$, $df = 117$, $p = .32$), although the mean for the cognitive demand of tasks linked to fiction texts ($\mu = 2.15$) was slightly higher than the mean for nonfiction texts ($\mu = 2.00$).

⁷ These effect sizes are not strictly comparable to the effect sizes for assignment cognitive demand because the comprehension and writing factor measures were standardized. Unlike the assignment measures, they do not have a meaningful natural metric. As a consequence, the observed range in our sample for the independent covariate for assignment cognitive demand ranges from 1 to 3.2, whereas the range for the standardized covariates is from -2 to 2 and thus contains four one-unit intervals as opposed to just over two.

REFERENCES

- Applebee, A.N. (1984). Writing and reasoning. *Review of Educational Research*, 54(4), 577–596. doi:10.3102/00346543054004577
- Applebee, A.N., & Langer, J.A. (2011). A snapshot of writing instruction in middle schools and high schools. *English Journal*, 100(6), 14–27.
- Applebee, A.N., & Langer, J.A. (2013). *Writing instruction that works: Proven methods for middle and high school classrooms*. New York, NY: Teachers College Press.
- Bangert-Drowns, R.L., Hurley, M.M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, 74(1), 29–58. doi:10.3102/00346543074001029
- Beck, I.L., & McKeown, M. (2001). Text talk: Capturing the benefits of read-aloud experiences for young children. *The Reading Teacher*, 55(1), 10–20.
- Beck, I.L., & McKeown, M. (2006). *Improving comprehension with Questioning the Author: A fresh and expanded view of a powerful approach*. New York, NY: Scholastic.

- Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education*, 13(3), 259–278. doi:10.1016/S0742-051X(96)00024-8
- Boscolo, P., & Carotti, L. (2003). Does writing contribute to improving high school students' approach to literature? *L1-Educational Studies in Language and Literature*, 3(3), 197–224. doi:10.1023/B:ESLL.0000003685.30925.c4
- Braker-Walters, B.A. (2014). Informational text and the Common Core: A content analysis of three basal reading programs. *SAGE Open*, 4(4). doi:10.1177/2158244014555119
- Business Roundtable. (2009). *Getting ahead—staying ahead: Helping America's workforce succeed in the 21st century*. Washington, DC: Author. Retrieved from businessroundtable.org/sites/default/files/BRT_Getting_Ahead_online_version_1.pdf
- Carbonaro, W.J., & Gamoran, A. (2002). The production of achievement inequality in high school English. *American Educational Research Journal*, 39(4), 801–827. doi:10.3102/00028312039004801
- Correnti, R. (2007). An empirical investigation of professional development effects on literacy instruction using daily logs. *Educational Evaluation and Policy Analysis*, 29(4), 262–295. doi:10.3102/0162373707309074
- Correnti, R., Matsumura, L.C., Hamilton, L.S., & Wang, E. (2012). Combining multiple measures of students' opportunities to develop analytic text-based writing. *Educational Assessment*, 17(2/3), 132–161. doi:10.1080/10627197.2012.717035
- Correnti, R., Matsumura, L.C., Hamilton, L.S., & Wang, E. (2013). Assessing students' skills at writing in response to texts. *The Elementary School Journal*, 114(2), 142–177. doi:10.1086/671936
- Correnti, R., & Rowan, B. (2007). Opening up the black box: Literacy instruction in schools participating in three comprehensive school reform programs. *American Educational Research Journal*, 44(2), 298–339. doi:10.3102/0002831207302501
- Correnti, R., Rowan, B., & Camburn, E. (2003, April). *School reform programs, literacy practices in third-grade classrooms, and instructional effects on student achievement: Notes from the first year of the Study of Instructional Improvement*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Crosson, A., Matsumura, L.C., Correnti, R., & Arlotta-Guerrero, A. (2012). Writing tasks and students' use of academic language. *The Elementary School Journal*, 112(3), 469–496. doi:10.1086/663298
- Doyle, W. (1983). Academic work. *Review of Educational Research*, 53(2), 159–199. doi:10.3102/00346543053002159
- Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. *Educational Psychologist*, 23(2), 167–180. doi:10.1207/s15326985ep2302_6
- Doyle, W., & Carter, K. (1984). Academic tasks in classrooms. *Curriculum Inquiry*, 14(2), 129–149. doi:10.2307/3202177
- Escher, A.L. (2015). *Constructing knowledge through writing: An analysis of writing tasks in 11th grade ELA textbooks* (Unpublished doctoral dissertation). Pittsburgh, PA: University of Pittsburgh.
- Fisher, D., Frey, N., & Lapp, D. (2012). *Text complexity: Raising rigor in reading*. Newark, DE: International Reading Association.
- Gilbert, J., & Graham, S. (2010). Teaching writing to elementary students in grades 4–6: A national survey. *The Elementary School Journal*, 110(4), 494–518. doi:10.1086/651193
- Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Washington, DC: Center for Educator Compensation Reform.
- Graham, S., Cappizi, A., Harris, K.R., Hebert, M., & Murphy, P. (2014). Teaching writing to middle school students: A national survey. *Reading & Writing*, 27(6), 1–28.
- Graham, S., & Harris, K.R. (1993). Self-regulated strategy development: Helping students with learning problems develop as writers. *The Elementary School Journal*, 94(2), 169–181. doi:10.1086/461758
- Graham, S., & Harris, K.R. (2006). Strategy instruction and the teaching of writing. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187–207). New York, NY: Guilford.
- Graham, S., & Hebert, M.A. (2010). *Writing to read: Evidence for how writing can improve reading*. Washington, DC: Alliance for Excellent Education.
- Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools*. Washington, DC: Alliance for Excellent Education.
- Harris, K.R., & Graham, S. (2009). Self-regulated strategy development in writing: Premises, evolution, and the future. *British Journal of Educational Psychology* [Monograph series], 2(6), 113–135.
- Hiebert, E.H., & Mesmer, H.A.E. (2013). Upping the ante of text complexity in the Common Core State Standards: Examining its potential impact on young readers. *Educational Researcher*, 42(1), 44–51. doi:10.3102/0013189X12459802
- Hillocks, G., Jr. (1984). What works in teaching composition: A meta-analysis of experimental treatment studies. *American Journal of Education*, 93(1), 133–170. doi:10.1086/443789
- Hillocks, G., Jr. (2006). Middle and high school composition. In P. Smagorinsky (Ed.), *Research on composition: Multiple perspectives on two decades of change* (pp. 48–77). New York, NY: Teachers College Press.
- Junker, B.W., Matsumura, L.C., Crosson, A., Wolf, M.K., Levison, A., Wiesberg, J., & Resnick, L. (2006). *Overview of the Instructional Quality Assessment* (CSE Technical Report No. 671). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California.
- Koretz, D., & Hamilton, L. (2006). Testing for accountability in K–12. In R.L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531–578). Westport, CT: American Council on Education.
- Langer, J.A., & Applebee, A.N. (1987). *How writing shapes thinking: A study of teaching and learning* (NCTE Research Report No. 22). Urbana, IL: National Council of Teachers of English.
- Marshall, J.D. (1987). The effects of writing on students' understanding of literary texts. *Research in the Teaching of English*, 21(1), 30–63.
- Maryland State Department of Education. (2014). *MSA Reading—BCR rubric—grades 3–8*. Retrieved from mdk12.org/assessments/k_8/readingscoring_bcr.html
- Matsumura, L.C. (2005). *Creating high-quality classroom assignments*. Lanham, MD: ScarecrowEducation.
- Matsumura, L.C., Garnier, H., Pascal, J., & Valdés, R. (2002). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment*, 8(3), 207–229. doi:10.1207/S15326977EA0803_01
- Matsumura, L.C., Garnier, H., Slater, S.C., & Boston, M.B. (2008). Measuring instructional interactions 'at-scale'. *Educational Assessment*, 13(4), 267–300. doi:10.1080/10627190802602541
- Matsumura, L.C., Patthey-Chavez, G.G., Valdés, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *The Elementary School Journal*, 103(1), 3–25. doi:10.1086/499713
- Monte-Sano, C., & De La Paz, S. (2012). Using writing tasks to elicit adolescents' historical reasoning. *Journal of Literacy Research*, 44(3), 273–299. doi:10.1177/1086296X12450445
- National Center for Education Statistics. (2012). *The Nation's Report Card: Writing 2011* (NCES 2012-470). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

- National Commission on Writing in America's Schools and Colleges. (2003). *The neglected "R": The need for a writing revolution*. Reston, VA: The College Board.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards in English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.
- Newell, G.E. (2006). Writing to learn: How alternative theories of school writing account for student performance. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 235–247). New York, NY: Guilford.
- 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(3), 273–304.
- Newkirk, T., & Atwell, N. (Eds.). (1982). *Understanding writing: Ways of observing, learning and teaching, K–8*. Chelmsford, MA: Northeast Regional Exchange.
- Newmann, F.M., Bryk, A.S., & Nagaoka, J. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence*. Chicago, IL: Consortium on Chicago School Research.
- Newmann, F.M., & Wehlage, G.G. (1993). Five standards of authentic instruction. *Educational Leadership*, 50(7), 8–12.
- O'Hallaron, C.L. (2014). Supporting fifth-grade ELLs' argumentative writing development. *Written Communication*, 31(3), 304–331. doi:10.1177/0741088314536524
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Raudenbush, S.W., Rowan, B., & Kang, S.J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to U.S. high-school data. *Journal of Educational and Behavioral Statistics*, 16(4), 295–330.
- Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *The Elementary School Journal*, 105(1), 75–101. doi:10.1086/428803
- Spillane, J.P., Reiser, B.J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387–431. doi:10.3102/00346543072003387
- Stein, M.K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, 2(1), 50–80. doi:10.1080/1380361960020103
- Voss, J.F., & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research*, 27(3), 255–265. doi:10.1016/S0883-0355(97)89733-9
- Wiley, J., & Voss, J.F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, 91(2), 301–311. doi:10.1037/0022-0663.91.2.301
- Yeh, S.S. (1998). Empowering education: Teaching argumentative writing to cultural minority middle-school students. *Research in the Teaching of English*, 33(1), 49–83.

Submitted November 13, 2014
 Final revision received May 28, 2015
 Accepted June 1, 2015

LINDSAY CLARE MATSUMURA (corresponding author) is an associate professor in the School of Education and a research scientist in the Learning Research and Development Center at the University of Pittsburgh, Pennsylvania, USA; e-mail lclare@pitt.edu.

RICHARD CORRENTI is an associate professor in the School of Education and a research scientist in the Learning Research and Development Center at the University of Pittsburgh, Pennsylvania, USA; e-mail rcorrent@pitt.edu.

ELAINE WANG is a doctoral candidate in the Learning Sciences and Policy program in the School of Education at the University of Pittsburgh, Pennsylvania, USA; e-mail elw51@pitt.edu.

Supporting Information

Additional supporting information may be found in the online version of this article on the publisher's website:

- Appendix A: RTA Scoring Rubric and Anchor Papers
- Appendix B: Effect of Assignment Cognitive Demand on Overall RTA Score