The Visual Thinking Strategies Approach to Teaching Argument Writing:

A Professional Development Model

The Elementary School Journal, 2020

Keith Smolkowski[1], Lisa A. Strycker[1], Lynne Anderson[2], Peggy Marconi[2], and Lisa Abia-Smith[3]

[1]Oregon Research Institute, Eugene OR USA

Keith Smolkowski: 541-484-2123; keiths@ori.org

Lisa A. Strycker: 541-484-2123; lisas@ori.org

[2]College of Education, University of Oregon, Eugene, OR USA

Lynne Anderson: 541-954-7415; lynneai@uoregon.edu

Peggy Marconi: 541-505-6987; pmarconi@uoregon.edu

[3]Jordan Schnitzer Museum of Art, University of Oregon, Eugene, OR USA

Lisa Abia-Smith: 541-346-0966; abia@uoregon.edu

Abstract

This pilot study evaluated a novel professional development (PD) model designed to prepare K-12 educators in high-need, rural school districts to teach evidence-based argument writing. The PD showed teachers how to engage students in discussions about visual art using Visual Thinking Strategies to enhance students' argumentation writing. Twenty intervention teachers completed the professional development program and conducted lessons in Visual Thinking Strategies and argument writing. Intervention teachers, relative to 34 comparison teachers, improved on their perceived knowledge of critical thinking concepts, ability to teach critical thinking concepts, knowledge of concepts related to argument writing and their actual ability to perceive and analyze artistic work and understand characteristics used to convey intent and meaning (Hedges's $g > 0.50$). Students in the intervention condition ($N = 210$ in K-11) improved their argument writing skills. Given the quasi-experimental design and small sample size, these findings should be viewed with caution but suggest promise.

Keywords: critical thinking, reasoning, argument writing, art, visual literacy

The Visual Thinking Strategies Approach to Teaching Argument Writing:

A Professional Development Model

The STELLAR Project (Strategies for Technology-Enhanced Learning and Literacy through ARt) aimed to address two growing challenges in schools: (a) the need to improve argument writing among students and (b) students' limited exposure to art. The mandate to improve students' argument writing stems in part from Common Core State Standards (CCSS), a nationwide state-level initiative intended to help students develop the skills needed to succeed in college and future careers (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). The CCSS for English Language Arts encourage the use of evidence and sources to support written arguments and analysis; these arguments can be "rhetorically and linguistically complex" (Olson, Scarcella, & Matuchniak, 2015, p. 578). In implementing the CCSS, many school districts have increasingly focused on core subjects, such as math and science, and have reduced or eliminated funding for other subjects such as art (Hambek, 2016). As fewer schools—especially those in rural or high-need districts—provide regular access to art specialists, students have fewer opportunities to develop visual literacy skills (Parsad & Spiegelman, 2012) and meet the National Core Arts Standards intended to shape instruction and student achievement in visual and other arts (National Coalition for Core Arts Standards, 2014). In the educational context, the term *visual literacy* describes the ability to interpret, recognize, appreciate, and understand information presented through visual images (Finley, 2014).

The novel intervention approach of the STELLAR Project sought to explicitly tie visual literacy with argument writing. Through an innovative professional development (PD) program, the STELLAR Project aimed to show teachers how to engage students in discussions about

visual art in order to enhance students' argumentation writing skills. Specifically, the STELLAR PD was designed to help teachers (a) increase art exposure, enhance art appreciation, and increase visual literacy for themselves and their students; (b) scaffold students' critical thinking and argument writing—especially the use of supporting evidence—through visual literacy; and (c) integrate visual literacy, critical thinking, and argument writing across the curriculum. The purpose of this manuscript is to report on the development and implementation of the STELLAR Project intervention, and its effects on teachers' perceived and actual knowledge and skills related to teaching argument writing and visual literacy.

Argumentation

The term "argument" is often used to refer to a quarrel, heated debate, or exchange of opposing views. In the context of critical thinking, argument may be more narrowly defined as a set of connected statements supporting the truth of a claim (Hillocks, 2011). The strength of the argument depends on its premises and their links to the claim; thus, it may be said that "arguments give us reasons to believe a claim is true" (Cook, Ellerton, & Kinkead, 2018, p. 2). Argument is essential to a functioning society. It underlies politics, ethics, management, science, decision making, negotiation, and much more. Students who learn how to understand and interpret arguments put forth by others, and to construct strong arguments themselves, can better grasp complicated content and assess the merits of various positions. According to Gilbert (1996), "One major reason for arguing, then, is to learn: to explore, probe, and test a belief or point of view" (p. 9).

Parts of an argument. Argumentation instruction is often based on the structure advanced by Stephen Toulmin (1958/2003; Klein & Rose, 2010). Within this framework, an argument consists of a claim, facts, and warrants—sometimes with additional backing, rebuttals,

and qualifiers. A claim (or thesis, or conclusion) is a proposition which asserts a clear and unambiguous point of view. Support for a claim comes from evidence (or grounds, data, facts, or reasons). Evidence is fundamental to arguments. Hillocks (2011) contends that the construction of most arguments should begin with data. An argument is usually more convincing if facts (used as evidence) are linked to the claim through a warrant (Toulmin, 1958/2003), which describes how evidence validates the claim. Warrants make logical, general inferences; they may be questioned as to their acceptability, applicability, or correctness. In casual conversation, warrants are often unstated. But this practice hides the logic of an argument, and for this reason warrants in written arguments should be expressed. By clearly stating warrants, argument makers can identify and address potential flaws; explicit warrants also allow the audience to follow the reasoning and ask critical questions. As hypothetical statements, warrants may require their own evidence or backing, which is another type of fact. Evidence lays the foundation of an argument, but most facts in an argument lend authority to warrants through backing (Toulmin, 1958/2003).

Teaching argumentation. Making an argument—even an apparently simple one—can quickly become complex (Verlaan, Ortlieb, & Verlaan, 2014). Teaching argumentation can be challenging, too, especially for teachers who lack professional preparation and are "unaware of [the] concepts" (Newell, Beach, Smith, & VanDerHeide, 2011, p. 277) of argumentation. A further complication to teaching argumentation is that teachers may underestimate how much background knowledge students require (Newell et al., 2011) in order to make an argument. For example, few high school students—let alone middle or elementary school students—possess a deep understanding of climate science. Students learning how to write an argument can easily become overwhelmed if they must simultaneously acquire a considerable load of content knowledge while stretching their language skills.

Many teachers forgo argumentation instruction altogether in an effort to minimize conflict in their classrooms. This practice, though understandable, stifles constructive disagreement. It also fails to recognize the need to explicitly teach students how to construct an argument and how to evaluate the arguments of others. Lacking instruction in argumentation, students often "see no need to present evidence or show why it is relevant" (Hillocks, 2011, p. 15). Teachers can teach argumentation effectively, encourage free discussion, and limit conflict by focusing on the concepts and procedures of argumentation rather than on subject matter. Because students may be "reluctant to adopt stances that their teacher may not endorse (Beck, 2006)" (Newell et al., 2011, p. 277), contentious topics may be avoided. Visual Thinking Strategies (VTS) allows teachers to focus on concepts of argumentation without the need for substantial background knowledge or controversial topics.

Visual Thinking Strategies

VTS was developed by Phillip Yenawine at the Museum of Modern Art in New York City (Yenawine, 2013). VTS is an instructional method that helps teachers engage their students in discussions about visual art—and in the process enhance students' critical thinking, language, and visual literacy skills. VTS has been shown to be an appropriate inquiry-based teaching strategy for all grade levels (Yenawine, 2013), not only promoting the development of art literacy but also improving student understanding and appreciation of course material in other areas, such as building background knowledge and fostering critical thinking in close reading in elementary school (Cappello & Walker, 2016); strengthening observation, critical thinking, and assessment skills in nursing (Moorman, Hensel, Decker, & Busby, 2017; Nanavaty, 2018); and learning clinical skills in medical school (Karwowski, Miller, & Sarfaty, 2014).

During a teacher-led VTS session, students are asked to view a piece of art and answer three questions about it: (a) "What is going on in this picture?" (b) "What do you see that makes you say that?" and (c) "What else can we find?" Each question is followed by student discussion. The first question asks students to think about and interpret what they see—the action or story in the image—leading to statements that reflects their interpretation. In the terms of argumentation, these statements can be viewed as *claims*. The second question asks students to share what they saw in the picture that led to their interpretation. This question asks the student to provide *evidence* for the claim and link that evidence to the claim with a *warrant*. Because students rarely exhaust all possibilities, the third question asks students to think more about the picture, which often leads student to express a different or expanded interpretation, a new claim. The teacher then requests observed details or evidence to support the new interpretation.

VTS and Argument Writing

Some teachers have taught argument writing for decades, others have employed VTS to teach aesthetic development in schools (e.g., Housen & Yenawine, 2000), and many have drawn upon VTS (2020) to teach writing. Although VTS also holds promise as a method for teaching argumentation, no research prior to the STELLAR Project, to our knowledge, has explicitly tied VTS to argument writing. By its nature, visual art is compelling, complex, and ambiguous. It inspires opinions from viewers, elicits differing viewpoints from different viewers, and requires no special background or technical knowledge to appreciate. In a classroom VTS session, students naturally engage in the fundamentals of argumentation when they answer their teacher's three questions about a piece of visual artwork. They (a) make a claim, (b) specify evidence and warrants, and may also provide (c) additional backing, (d) rebuttals, or (c) qualifiers. VTS naturally leads to open discussions of how opinions are supported by evidence—the basis of

argumentation. In the relatively neutral context and safe space of visual art and VTS, students become empowered to develop arguments that may differ from their classmates or teacher (Yeh, 1998).

Once students have learned, through VTS, to make practical arguments backed with evidence, teachers can then direct those skills into writing an argument. The VTS lesson begins by asking students to take a silent moment to look at an image and interpret what they see. In the same way, when taking a position or making argument about an issue, students are encouraged to pause and think before entering the conversation. With VTS, students learn to move beyond knee-jerk responses. Taking time to look at an image, passage, or video spurs their initial thinking and leads to making a well-supported, reasoned argument. This process underscores how a writer thinks and ultimately how a writer puts that thinking on paper.

As an example, students might engage in a group VTS session about an image of "Fall" by Wendy Red Star—a self-portrait photograph of a Native American woman surrounded by artificial props and backdropped by an inflatable deer and a painted Western autumnal landscape. In the classroom discussion, the teacher asks the three VTS questions about the image, and students give oral answers. Afterwards, students are assigned to compose a written argument. A fifth-grade student might write about this image: "I see that a bull died there because there is a remaining bull skull. I see that the girl in the picture is maybe a Native American because she's wearing Native American clothes. I see that it's fall because there are dead leaves. I see she has an animal because there's an animal right next to her and it's staying calm. I see she's on the mountain because in the background the mountain is as high as her." Without VTS instruction, the fifth grader's writing sample would likely lack the key elements of argumentation, such as "I see leaves, a doe, flowers, and other stuff." This same image can be used with younger students

to help them form arguments—which are likely to be less-nuanced, and perhaps given orally in the lower grades. After a VTS session, a kindergarten student might say about this image, "A girl is sitting down in a fake store. I think it is fake because there's a fake background behind her, with a fake tree and a fake deer, and there's a weird skull right there of some animal." Without VTS, the answer would probably not contain a claim or evidence: "I see a person. I see a tree. I see leaves."

The STELLAR Project

The STELLAR Project was designed specifically to create a culture of argumentation in classrooms through exposure to the visual arts using VTS. The project's design and implementation combined the interests and expertise of its two lead organizations, the Oregon Writing Project at the University of Oregon (an affiliate of the National Writing Project), experienced in providing high quality professional development on writing to rural teachers, and the University of Oregon's Jordan Schnitzer Museum of Art, with experience using VTS to teach visual literacy skills to teachers and students. Funded by the U.S. Department of Education, the STELLAR Project was a collaborative effort in which researchers, writing experts, VTS professionals, and school districts worked together to help teachers in rural school districts enhance classroom instruction in argument writing and the visual arts. The project's funding mechanism required that the intervention be carried out districtwide, in small, rural communities, with the goal of building instructional capacity for evidence-based reasoning and argument writing using VTS. For this reason, the intervention covered all K-12 teachers in all subject areas in the recruited districts.

The theory of change underlying the STELLAR Project posited that (a) an intensive and sustained PD program will impact the arts knowledge and skill of rural teachers; (b) engaging in

this type of PD program will improve teachers' knowledge and skill in VTS, evidence-based reasoning, and argument writing, quantified by measures of teachers' critical thinking, argument writing, and visual literacy; and (c) gains in teacher knowledge and skill will positively impact students' visual literacy, critical thinking skills, and argument writing, quantified in part by measures of student argument writing. The present study tested the two hypotheses implicit in the first two parts of our theory of change (a and b) regarding teachers' engagement in PD and change in knowledge and skills. Some student data also were collected, but student outcomes were not the primary focus.

The project created an intensive and sustained PD program to show teachers how to use VTS to promote their students' critical thinking, visual literacy, argumentation, and ultimately argument writing skills. The PD program gave teachers a practical introduction to argumentation instruction. It showed teachers how to lead VTS sessions in their classrooms, with the goal of improving argument writing skills among their students. The PD program followed Toulmin's framework, interpreted through Hillocks (2011), with modifications as needed. For example, elementary school teachers who found the distinction between grounds and backing too complex for their students could still teach the basic principles of producing a claim, supporting it with evidence, and bridging the two with a warrant. PD activities encouraged the use of VTS and argumentation across subject domains.

The STELLAR PD program was developed in conjunction with teachers over 2 years. The project team blended theories of instruction and argumentation as teachers used VTS to enhance visual literacy and teach argumentation to their students. This iterative, participatory development process, a model recommended for successful school-based implementations (Anderson, Marconi, Smolkowski, & Strycker, 2019), was intended to ensure that the final PD

program was acceptable and useful to teachers and that teachers could feasibly and successfully implement VTS with argument writing in their classrooms.

The PD program assumed that (a) teachers could learn strategies for interpreting visual images that would improve their ability to look closely and make sense out of what they see, and teach these strategies to their students; (b) the ability to interpret images and other visual stimuli would transfer from art to other domains, such as reading, science, and social studies; (c) the production of claims and evidence during VTS sessions would aid teachers in teaching argument writing across domains; and (d) intensive effort to increase and refine visual literacy, critical thinking, and argument writing skills would have a positive impact on student outcomes across subject areas. The STELLAR Project provided teachers with initial and ongoing PD, as well as peer coaching, in line with recommendations by Fixsen, Naoom, Blasé, Friedman, and Wallace (2005).

Research Questions

The main purpose of this preliminary project was to develop the PD intervention, and then provide an initial test of its feasibility and influence on teachers. Teacher measures created for the project were designed to document the effects of the PD program on the areas it targeted: teachers' perceived knowledge and skills, and actual ability, to understand and teach critical thinking skills with VTS to students. Despite the focus on teachers, basic measures of student argument writing were collected. We addressed one primary and one secondary research question. First, to what extent did the PD program improve teachers' perceived knowledge and skills related to VTS, argumentation, and argument writing, and increase their actual ability to teach these strategies? Second, to what extent did the students of trained teachers improve argument writing skills?

Method

This study was conducted in the Pacific Northwest in the 2017–2018 academic year with a sample of K-11 teachers and students. The design compared two groups of teachers, an immediate-intervention group and a waitlist comparison group. Because the intervention was intended to be districtwide, project staff assigned teachers to condition by district, matching district characteristics as much as possible to balance the two groups. Although nonrandom, groups did not significantly differ on baseline measures, indicating no systematic difference prior to treatment assignment. This quasi-experimental comparison group design is appropriate for small pilot studies like this, which aim to gather evidence of feasibility prior to proposing a more rigorous randomized controlled trial with a control group (Kraemer, Mintz, Noda, Tinklenberg, & Yesavage, 2006). The design is useful for evaluating changes in teacher knowledge and skills as a function of the project's PD program—the primary focus of this study—and for assessing the promise of impact on student outcomes, but causal conclusions must be drawn carefully. Assessments included teacher measures from both groups, as well as student writing samples from the immediate-intervention group before and after teacher training.

Sample

Participants were recruited via fliers and announcements at high-need rural school districts. Intervention and comparison group teachers were recruited from different districts in the same counties. Characteristics of participating school districts are presented in Table 1.

Recruitment procedures and participation requirements were identical for teachers in both conditions, except that intervention teachers completed the PD program in early winter of the first school year while comparison group teachers taught their children as they normally do during the first year. All teachers in comparison districts were offered an abbreviated version of

the PD training in the fall of the next school year, after the completion of all study activities. The PD for teachers in the waitlist comparison group was voluntary, and no data were collected from comparison group teachers or their students after they completed the PD.

Data were collected from 20 intervention group teachers and 34 comparison group teachers from all grade levels and from a variety of subject areas before and after the PD for the intervention group. Demographic data were not collected for teachers, but all teachers were recruited from rural districts of similar sizes within the same two counties. There were no baseline differences between the two teacher groups on the two primary measures, described below. Informed consent was obtained for all participating teachers. Seventeen of the intervention group teachers provided pre- and post-intervention writing samples from students (mean = ~12 students/teacher). Student data were anonymous; teachers were shown how to use codes to track students' writing samples while removing all identifying information.

The study gathered anonymous writing samples in intervention districts from 210 students anonymously: 206 students at pretest, 206 at posttest, and 202 at both time points. Student writing samples were from a range of grades intended to represent districtwide implementation: Kindergarten (n = 12; 5.7%), Grade 2 (n = 59; 28.1%), Grade 4 (n = 36; 17.1%), Grade 5 (n = 12; 5.7%), Grade 7 (n = 17; 8.1%), Grade 8 (n = 36; 17.1%), Grade 9 (n = 25; 11.9%), and Grade 11 (n = 13; 6.2%). As the students were anonymous, demographic data were not collected.

Procedures

STELLAR PD. The aim of the STELLAR PD was to give teachers practical instruction and support in conducting VTS in their classrooms. Teachers received step-by-step guidance in how to choose appropriate works of visual art for VTS, facilitate classroom discussions of

artwork using the three questions of VTS, help students in these discussions learn and practice argumentation skills, and direct students in applying these skills to argument writing (Abia-Smith, Smolkowski, Strycker, Anderson, & Marconi, 2020).

All participating intervention teachers received initial instruction in visual literacy and training to implement VTS through an intensive full-day institute in January. Participants were given an overview of the project and instruction on VTS delivered by a VTS trainer external to the project. Training included an introduction to VTS principles, discussion and reflection on visual literacy, instruction and discussion on the specific components of VTS, and practice using VTS in the Jordan Schnitzer Museum of Art galleries at the University of Oregon. Teachers also learned how they might adapt VTS for different content areas. For instance, a science teacher could use VTS for a lab experiment, a math teacher could use VTS for a story problem, a language arts teacher could use VTS to interpret a poem, a Spanish teacher could use VTS to view photographs of Spanish culture, and a history teacher could use VTS with historical newspaper images. Teachers understood that the VTS approach represented a set of strategies rather than a graded curriculum. They were encouraged to decide for themselves what worked for their classes and what level of complexity was appropriate for their students. The workshop ended with a discussion of how to introduce VTS into school curricula, an overview of web resources and additional PD opportunities, and a review of critical thinking assessment.

A second, half-day training session was used to introduce teachers to a culture of argumentation and the components of argument writing. Teachers learned simple strategies for teaching argument writing, the connections between VTS and argument writing, and how to modify their instruction to accommodate differences in student needs across grade levels. This enabled them to teach the components of argumentation and argument writing at different levels

of complexity. The half-day workshop also introduced teachers to *Second Life*, an online, immersive virtual environment developed by the project so teachers from rural districts could attend regular follow-up training sessions, share their successful practices with colleagues, and work collaboratively on curriculum design and lesson planning. During the training session, teachers were enrolled in *Second Life*, acquired and personalized their avatars, and learned how to navigate in *Second Life* so they could attend follow-up PD activities from their homes.

Over a period of 5 months, intervention teachers participated in online PD via *Second Life* for about 2 hours every 2 weeks. They learned to extend and refine their skills in teaching with VTS, integrate VTS into their curricula, and teach critical thinking and argument writing. This was an innovative use of *Second Life* for rural teachers in primary and secondary schools. *Second Life* supplied a three-dimensional immersive venue for teacher meetings, workshops, presentations, and collaboration. Virtual meetings in *Second Life* enabled frequent and ongoing PD, as well as collaborations across districts, despite rural teachers' dispersed geographic locations. The use of social networks and virtual worlds for PD has been described as transformative for those who have experienced its power to promote professional communities online and support peer-to-peer learning. Example *Second Life* PD topics included (a) elements of VTS, (b) practicing VTS with art images, (c) refining VTS paraphrasing, (d) VTS and journaling, (e) argument in elementary classrooms, (f) VTS and elements of argument, (g) supporting students thinking through paraphrasing, and (h) writing the argument essay. Teachers also learned how to teach argument writing aligned with the CCSS. The *Second Life* activities, combined with the two in-person workshops, provided intervention teachers with 40+ hours of PD. The PD focused on VTS and argument writing, rather than writing instruction *per se*, under the assumption that in-service teachers already had been trained to teach writing. Intervention

teachers were provided with copies of three books to specifically enhance their knowledge of teaching argument writing and VTS: *Teaching Argument Writing, Grades 6–12: Supporting Claims with Relevant Evidence and Clear Reasoning* (Hillocks, 2011), *They Say / I Say: The Moves That Matter in Academic Writing* (3rd ed., Graff & Birkenstein, 2014), and *Visual Thinking Strategies Basic Manual Grades 3–5* (Housen & Yenawine, 2000).

The STELLAR Project activities were similar to teacher study groups or professional learning communities, which have been used for teacher PD (e.g., Jayanthi et al., 2018). Such activities, however, are often difficult for teachers in rural districts. *Second Life* allowed teachers to meet regularly despite their disparate geographic locations, and evening sessions permitted teachers to attend after school from home. The online venue avoided the costs of food, travel, and lodging associated with most PD provided by universities and other organizations for rural districts. *Second Life* sessions were also captured on video so that teachers who missed or wanted to review a session could watch asynchronously. Attendance and participation in the *Second Life* follow-up sessions were high, and the sense of community that developed was long-lasting. Teachers reported that the guests and presenters were fascinating, and the conversations were rich. Many participants still count the teachers in this project as treasured colleagues.

Teachers were also offered opportunities for mentoring by colleagues (mentor teachers) and on-site support from project staff or visiting artists. Most classrooms participated in one class field trip to the Jordan Schnitzer Museum of Art, where teachers and students practiced VTS and honed critical thinking skills with the support of museum docents trained in VTS.

All elements of the PD were standardized so that they were delivered exactly the same way to all participating teachers. Participants also completed surveys and tests of visual literacy and argument writing instruction. Teachers were encouraged to review and discuss (anonymous)

student writing samples with trainers to learn how to evaluate their students' writing, help refine their approach to instruction in argument writing, and share successes. These measures were used to evaluate the project and are detailed below.

Teachers in the waitlist comparison group were introduced to the *Second Life* digital program and encouraged to use it for follow-up training, sharing with colleagues, collaborating with other teachers, and lesson planning. Similar to intervention teachers, comparison teachers received copies of three books about VTS and argument writing, as well as opportunities for mentoring (including opportunities to discuss with mentors how to evaluate students' writing). The study did not conduct any formal follow-up with these teachers.

Classroom implementation. Project staff members did not directly teach students in classrooms. Participating teachers provided classroom instruction in VTS, critical thinking, and argument writing to students based on the PD they received. Teachers used images from the VTS Image Library (subscription provided through the project) or other sources, objects, documents, infographics, text, photos, puzzles, data displays, and problems. Most used digital tools or online resources. In the winter and spring of the school year, teachers were expected to complete at least nine VTS and argument writing sessions in their classrooms—a total of 4.5 to 18 hours of instruction, assuming 30 minutes to 2 hours per lesson.

Implementation fidelity. Classroom implementation of the VTS activities was measured by teachers' electronic logs (see STELLAR eLogs section, below**)** and monitored informally. Teachers provided information through online logs and were observed by project staff when implementing VTS in the museum galleries during field trips. These were coupled with discussions and coaching by project staff members. Project staff members and mentor teachers worked with participating teachers to motivate implementation of VTS with fidelity in their

classrooms and to support the integration of VTS with argument writing. Mentor teachers were considered an important part of building capacity at the school district. They were chosen by the project staff and received additional training in how to continue to support other teachers in VTS and in argument writing.

Measures

Validated measures documenting teachers' critical thinking, argument writing, and visual literacy, and their ability to teach those skills were lacking. This presented a measurement challenge, which necessitated the development of two new instruments specifically for this project: the STELLAR Teacher Survey of Critical Thinking and Argument Writing (STELLAR Teacher Survey) and the STELLAR Test of Visual Literacy for Teachers (STELLAR Test for Teachers). Items for the STELLAR Teacher Survey were designed to assess constructs covered in the STELLAR PD. Items for the STELLAR Test for Teachers were based on the National Core Arts Standards for Visual Arts (www.nationalartsstandards.org; State Education Agency Directors of Arts Education, 2014). The STELLAR Teacher Survey and STELLAR Test for Teachers were administered to intervention and comparison group teachers at the same two time points. The pretest was given before the intervention group began the PD and the posttest was given at the end of the school year after they had finished classroom VTS implementations. The comparison group teachers then received an abbreviated version of the PD and were encouraged to use VTS in their classrooms but were not further assessed. Among intervention teachers, we also collected STELLAR VTS Teacher eLogs (STELLAR eLogs) and student argument writing samples. Data from teachers were collected online via Qualtrics survey software. Although the STELLAR Test for Teachers assessed several aspects of visual arts, the project did not provide explicit training on most of those topics. Hence, with the exception of one scale on this test

(perceive and analyze artistic work), and possibly a second (interpret intent and meaning in artistic work), this measure was considered an exploratory, supplemental measure.

STELLAR Teacher Survey. The STELLAR Teacher Survey asked teachers to rate their knowledge or expertise on a 5-point Likert-type scale. The survey included items used to create four scales related to argument writing and critical thinking: perceived knowledge of critical thinking concepts (7 items), perceived ability to teach critical thinking concepts (6 items), perceived knowledge of concepts related to argument writing (7 items), and perceived ability to teach argument writing (12 items). Item scores ranged from 1 to 5. Scales were computed based on the average of multiple items related to each construct and produced acceptable reliability estimates (Table 2; coefficient alphas = .88–.97).

STELLAR Test for Teachers. The STELLAR Test for Teachers also consisted of open-ended questions, each aligned with one of five anchor standards of the National Core Arts Standards for Visual Arts (State Education Agency Directors of Arts Education, 2014): three in Responding, one in Connecting, and one in Presenting. The items corresponded to the following anchor standards: organize and develop artistic ideas and work (AS 2), perceive and analyze artistic work (AS 7), interpret intent and meaning in artistic work (AS 8), apply criteria to evaluate artistic work (AS 9), and relate artistic work to cultural and historical contexts (AS 11). All of these items were included in the STELLAR Test for Teachers in order to cover a range of National Core Arts Standards for Visual Arts, but the STELLAR PD specifically addressed AS 7 and, to a lesser extent, AS 8. Questions about additional standards were added at the request of the funding agency.

The section on AS 7 included four open-ended questions that asked teachers to respond to two pieces of art with the prompts, "What's going on in this painting?," "List any cultural

observations you can make. . . ," and "Describe the individuals. . . ." Responses were scored on the quality of their observations, interpretations, and use of evidence. All samples were scored by five project staff readers with experience teaching VTS and argument writing, using the STELLAR protocol and blinded to condition and assessment time (pretest or posttest). For three of the four AS 7 questions, readers scored teachers' responses on (a) relevant observations (b) relevant interpretation, and (c) use of evidence to support interpretation, with each dimension scored from 0 to 3, for a total of nine scores. For a fourth question, which asked teachers to compare two written interpretations of the same piece of art, teachers received one score ranging from 0 (no relevant comparison) to 6 (multiple comparisons supported with evidence). The 10 AS 7 scores were summed to create a total score with a possible range of 0 to 33.

The section on AS 8 included five open-ended questions about how the artist used "five characteristics used to portray ideas and convey mood: (a) form and structure, (b) contextual information, (c) subject matter, (d) visual elements, and (d) use of media." Although these characteristics were not taught explicitly in the PD, like AS 7 teachers were scored on teachers' observations, interpretation, and use of evidence (0 to 3 each), concepts associated with both VTS and argumentation. The 0-to-3 scores for each of the three dimensions across the five questions (15 total scores) created a total score with a possible range of 0 to 45.

The remaining anchor standards were scored similarly, which produced acceptable reliabilities with coefficient alphas from .87 to .91. Eight teacher tests were randomly selected to check for inter-rater reliability, but no differences beyond 1 point were noted. This represents about 9% of the total 88 teacher pretest and posttest tests scored (8/88 = 9%).

STELLAR eLogs. As a measure of the extent to which teachers implemented the classroom intervention with fidelity, intervention group teachers were asked to complete

electronic logs describing their lessons incorporating VTS and argument writing. For each lesson, teachers were expected to fill out an eLog to document the subject areas covered, where the lessons were conducted, what images were used, level of student engagement and participation, and skills students exhibited. Completion of at least nine VTS and argument writing sessions (documented by eLogs) was considered strong evidence of classroom implementation fidelity.

Student writing samples. Teachers in the intervention group collected student writing samples but the project did not have sufficient resources to collect and score writing samples from comparison classrooms. For this assessment, teachers showed students an image and asked them to answer three questions about it: "What is going on in this picture?," "What do you see that makes you say that?," and "What more can you find?" Kindergarten and first-grade students could give oral answers that teachers wrote down; by second grade, students were expected to write their own answers. Project staff members then collected writing samples from teachers.

The sample for the present study was selected using the following procedure: (a) each student's writing samples from participating classroom teachers were matched across pretest and posttest, based on identification numbers assigned by the teacher, to form a pair; (b) the pairs were assigned numbers (e.g., 18a for pretest and 18b for posttest); (c) 12 pairs were randomly selected from each teacher's class using a random number within strata based on their initial writing quality (low, middle, and high); and (d) samples were shuffled by thoroughly mixing pretest and posttest before scoring them. To score pretest and posttest student argument writing samples, the project used the National Writing Project's (n.d.) *Using Sources Tool*, which does not differentiate scoring by grade (see scoring rubric in Appendix A). This tool includes scales rating the extent to which students use evidence from sources in their writing. Student work was

scored using the tool's established rubric on six dimensions: (a) word count, (b) presenting a claim, (c) including source material with multiple perspectives, (d) connecting source material to the claim, (e) organization contributing to argument development, and (e) language contributing to argument development. Samples were scored by five readers experienced with VTS and argument writing instruction using the STELLAR protocol and blinded to time point. All scales other than word count were scored as follows: 1 = *not present*, 2 = *developing*, 3 = *competently*, and 4 = *effectively*. An overall writing scale score was calculated based on the average of the five rating subscale scores. The scale produced acceptable reliabilities (coefficient alpha = .92 at pretest and .91 at posttest). Thirty student writing samples were randomly selected to check for inter-rater reliability, which represents about 7% of the total 413 student pretest and posttest writing samples scored. No differences beyond 1 point were noted.

Analyses

Descriptive statistics documented teacher and student outcome and ensured that the overall distributions reasonably met normality assumptions of the tests employed.

Teacher outcomes. We assessed intervention effects on each of the primary outcomes with a Time × Condition analysis with all available data. The analysis tests net differences between conditions on gains from pretest to posttest for individual teachers' outcomes. This test of net differences, sometimes called a *difference in differences* (DD) test, provides an unbiased and straightforward interpretation of the results recommended for quasi-experimental trials (Allison, 1990; Van Breukelen, 2006). The statistical model includes Time, Condition, and a Time × Condition interaction, with Time coded 0 at pretest and 1 at posttest and Condition coded 0 for control and 1 for intervention. The approach includes teachers even when missing either the pretest or posttest survey, but not both, as recommended by Graham (2009). With 41 teachers,

tests of Time × Condition, the critical DD test, used 39 degrees of freedom (df). Because the analysis included cases with partially missing data, the model-estimated DD does not precisely match differences in raw means.

Descriptive statistics of STELLAR eLog data were used to document intervention group teachers' classroom VTS lessons; eLog comments provided qualitative feedback.

Student outcomes. To test pretest-to-posttest gains in students' seven argument writing scores, separate paired t tests were performed. Due to the broad range of grade levels, gain scores allow for more interpretable results because they subtract out much of the classroom variability present in pretest and posttest scores. Because students were clustered by teacher, and student scores might be assumed to vary from teacher to teacher, seven corresponding two-level random-intercept/random-gain regression models (Level 1 = student; Level 2 = teacher) with maximum likelihood estimation also were conducted (Muthén & Muthén, 2017), using data from 201 students. The purpose of the multilevel models was to test for and account for any teacher clustering effects. These multilevel models estimated a random intercept and random gain (where possible) from pretest to posttest, allowing gains to vary by teacher. With full information maximum likelihood estimation, the analysis included students even when they were missing either pretest or posttest data.

Intraclass correlation coefficients (ICCs) were estimated to measure the similarity of student results from the same teacher. Pretest assessments generally account for variation across classrooms, but gain scores tend to have the higher-level influences of classrooms and schools subtracted out. With a wide range of grade levels, we expected large ICCs for pretests, which averaged .34 (range: .23–57) and indicated substantial similarity of students within classrooms

or, alternatively, substantial variability between classrooms. The ICCs for gains were much smaller, as expected, and averaged .04 (.00–.11).

Reporting and effect sizes. We computed Hedges' g values as an effect size to present differences between conditions in terms of posttest standard deviations (What Works Clearinghouse, 2017). Consistent with best practice recommendations of the American Statistical Association (Wasserstein & Lazar, 2016; Wasserstein, Schirm, & Lazar, 2019), we present 95% confidence intervals (CIs) for raw effects and Hedges' g values and focus on medium or larger effect sizes ($g > 0.50$) rather than statistical significance.

Results

Research Question 1: Teacher Results

STELLAR Teacher Survey. Table 2 provides descriptive statistics of the STELLAR Teacher Survey, split by intervention condition. There was little evidence to suggest differences between conditions at pretest on any scales (all $p > .05$), and for all scales, pretest differences favored the comparison condition.

The intervention group achieved greater gains relative to the comparison group for survey measures targeted by the intervention: perceived knowledge of critical thinking concepts (difference in differences [DD] = 0.81, 95% CI [0.37, 1.25], g = 1.35 [0.62, 2.09], t_{39} = 3.73, p = .0006), perceived ability to teach critical thinking concepts (DD = 0.56 [0.06, 1.05], g = 0.69 [0.08, 1.31], t_{39} = 2.28, p = .0281), and perceived knowledge of concepts related to argument writing (DD = 0.85 [0.31, 1.39], g = 1.05 [0.38, 1.72], t_{39} = 3.16, p = .0030). Although perceived ability to teach argument writing mean scores for intervention group teachers increased more than for comparison group teachers, the difference was more modest (DD = 0.32 [-0.13, 0.76], g

= 0.40 [-0.16, 0.96], t_{39} = 1.45, p = .1561). Overall, effect sizes for change in teachers' perceived

knowledge and skills ranged from medium to very large.

STELLAR Test for Teachers. The STELLAR Project focused on primarily on how well

teachers actually perceive and analyze artistic work (AS 7) and secondarily on their perceptions

and supporting evidence for the five characteristics used to convey intent and meaning (AS 8).

For AS 7, on average, comparison teachers made modest gains from 18.54 (SD = 8.36) to 19.87

(7.39) while intervention teachers' improvements were considerably larger, from 18.05 (8.42) to

25.59 (5.65). Group means did not differ at pretest (all p > .05), but the difference in gains was

large (DD = 6.16 [-0.52, 12.84], g = 0.92 [-0.08, 1.91], t_{39} = 1.87, p = .0695). For AS 8,

comparison teacher scores changed from 26.21 (10.35) to 26.96 (11.13) while intervention

teachers improved from 23.25 (11.76) to 31.65 (13.13), which showed considerable promise (DD

= 7.89 [0.17, 15.62], g = 0.66 [0.01, 1.30], t_{39} = 2.07, p = .0454). The two effect sizes from the

STELLAR Test for Teachers indicate important improvements in intervention teachers' actual

knowledge and skills. The magnitude was similar to the gains in teachers' perceived knowledge

and skills measured by the STELLAR Teacher Survey. Intervention and comparison teachers did

not differ substantially on other aspects of STELLAR Teacher Survey (g < 0.40), which were not

covered by the PD.

STELLAR eLogs. Fourteen intervention group teachers (from grades 2, 4, 5, 7, 8, 9, and

11) provided eLog data documenting VTS lessons administered from January 22, 2018 to June 8,

2018. The mean number of VTS lessons per teacher was 9.8 (range = 4–16), which indicated

good overall fidelity to the intervention protocol of at least nine VTS classroom sessions. Nearly

all of these VTS lessons occurred in the classroom; however, most teachers also accompanied

their students on one museum field trip with a VTS session in the galleries. Because none of

these teachers had previously conducted VTS lessons, this represented a substantial increase in the number of classroom lessons per month integrating critical thinking and argument writing with visual thinking. The level of student engagement and participation during lessons (scored as 1 = low, 2 = moderate, and 3 = high) averaged 2.40 (SD = 0.60). Specific skills students exhibited during the lessons included drawing inferences (84% of lessons), supporting observations with evidence (90%), evidential reasoning (61%), explanatory reasoning (66%), speculative thinking (54%), considering multiple perspectives (46%), and evaluative thinking (33%). The eLog data indicated that teachers used images from the VTS website or VTS Image Library for about 43% of their classroom VTS sessions.

Teachers wrote a total of 139 comments on the eLogs. . Examples of comments include: "Students seem to enjoy the process and they look forward to VTS when they see it coming up." "The kids have gotten really comfortable with knowing what the VTS questions are and what is expected of their behavior." "I had a great deal of participation in the discussion from a class that is usually pretty quiet." "More students are beginning to justify their observations without being asked." "I was so impressed with how much growth I saw despite doing very limited writing projects with the VTS images." Via the eLogs, teachers identified several limitations of the VTS approach, noting that some of the visual art offerings did not seem to be engaging to students, and that student discussions were not always focused. Examples of comments included: "I showed the Salvador Dali image, but students did not take it seriously…so I felt this was a fail" and "I have noticed that some of the students are becoming less patient about listening to other students' thoughts and blurt their responses."

Research Question 2: Student Results

Table 3 presents the results of the student writing assessment. For intervention schools only, we used paired t tests to compare gains in several aspects of argumentation from student writing samples. Multilevel modeling analyses produced similar results, and thus are not duplicated here. All student writing variables significantly improved from pretest to posttest, as follows: word count (pretest $M = 57.18$ [$SD = 56.64$], gain $M = 23.9$, $t_{202} = 9.08$, $p < .001$); presenting a claim (pretest $M = 2.34$ [0.78], gain $M = 0.53$, $t_{202} = 9.17$, $p < .001$), including source material with multiple perspectives (pretest $M = 2.44$ [0.69], gain $M = 0.57$, $t_{202} = 10.27$, $p < .001$), connecting source material to the claim (pretest $M = 2.22$ [0.87], gain $M = 0.69$, $t_{202} = 10.54$, $p < .001$), organization contributing to argument development (pretest $M = 2.20$ [0.94], gain $M = 0.75$, $t_{202} = 11.11$, $p < .001$), language contributing to argument development (pretest $M = 2.29$ [0.82], gain $M = 0.56$, $t_{202} = 9.57$, $p < .001$); and overall writing mean (pretest $M = 2.36$ [0.71], gain $M = 0.62$, $t_{202} = 13.41$, $p < .001$).

Discussion

This paper provides early findings from a pilot study of a PD program targeting teacher knowledge and skills in using visual art to teach argument writing. The primary aim of the STELLAR Project was to improve teachers' instruction in argument writing by integrating VTS with argumentation. The study was primarily a PD project, focusing on teachers' perceptions of their instruction regarding critical thinking and writing. Preliminary results are consistent with the conceptual theory underlying the STELLAR Project, which posited that an intensive and sustained PD program would positively affect teachers' perceived and actual knowledge and skill in VTS and argument writing. Participating teachers and administrators generally viewed the project positively, as shown by brief project videos of one teacher's experience

(https://vimeo.com/387517589) and one superintendent's reflections (https://youtu.be/b-OUqfrZJDs). The findings also provide limited evidence that the PD intervention has a positive impact on students' evidence-based reasoning and argument writing abilities. Given the small sample and quasi-experimental design of this pilot study, these results should be viewed with caution but suggest promise.

Implications for Teacher PD and VTS

In response to Research Question 1, this research found that intervention teachers perceived that the PD prepared them to teach their students argument writing through VTS. Intervention group teachers, relative to a comparison group, improved their perceived and actual knowledge and skills explicitly targeted by the PD program: perceived knowledge of critical thinking concepts, perceived ability to teach critical thinking concepts, perceived knowledge of concepts related to argumentative writing, and actual ability to perceive and analyze artistic work and understand characteristics used to convey intent and meaning. The effect sizes represented by Hedges's g values exceeded 0.65 standard deviations for all of these measures. The finding that the intervention group outperformed the comparison group on these variables should not be surprising, given that only intervention teachers received the STELLAR PD intervention. It is possible that the PD itself helped teachers perform at a higher level, regardless of the subject taught in the PD. The two teacher groups, however, performed more similarly on aspects of visual arts not covered by the PD.

These results extend previous research by suggesting positive outcomes from PD, especially in the realm of visual literacy. This was expected, given that the PD program was specifically designed to follow evidence-based, high-quality standards (e.g., Desimone, 2009). Anecdotally, teachers reported that they used the *Second Life* digital platform for ongoing PD for

VTS and argument writing. This suggests that rural teachers benefited from distance learning and online support and collaboration with mentors and other teachers. Previous research has demonstrated that PD leads to gains in teacher content knowledge, but results have been mixed about whether teacher gains transfer to student gains (De Kramer et al., 2012) or not (Dash, De Kramer, O'Dwyer, Masters, & Russell, 2012). The current study provides evidence that a PD approach can be effective in enhancing the perceived and actual ability of teachers to promote students' visual literacy in the arts as well as argument writing.

This study also contributes to the growing field of VTS. To date, few peer-reviewed manuscripts have been published documenting the effects of VTS on educational outcomes. Nelson (2017) found nine peer-reviewed articles on VTS in museum education, nine in medicine, and eight in K-12 education. Because many of these published papers are qualitative and focused on art literacy, this manuscript contributes to the nascent literature in K-12 education by reporting quantitative and qualitative evidence of the positive effects of VTS on teacher and student outcomes related to argument writing and visual literacy.

Teacher comments suggested some challenges with the implementation of VTS in the classroom, such as the lack of engagement due to art selection and unfocused or impatient students. Selecting art appropriate for students is a difficult skill as different images appeal to students of different ages. This aspect of the PD could have been strengthened. The STELLAR Project did not, however, offer any PD focused on behavior management, which may have been valuable for classrooms in which students were less practiced at sharing in open, classwide discussion.

Implications for Students

In response to Research Question 2, results of the STELLAR Project provide evidence of promise; that is, students of trained teachers successfully learn critical thinking and argument writing skills. Although student-level outcomes were not the primary target of this project, the ultimate test of a teacher PD program is the degree to which it enhances student learning. We evaluated students of trained teachers to document argument writing gains from pretest to posttest, under the assumption that students generally have no reason to improve on the measures without external intervention. This assumption likely has been violated for students in school because they generally learn writing skills throughout the year. We can only show that students who learned from teachers completing the PD program made appreciable gains in argument writing, suggest that the gains were not necessarily expected from regular classroom instruction, and advise caution in interpreting these results in the absence of a comparison group.

The findings of this study are in line with prior research, which has demonstrated that students produce better analytic writing—including using evidence to support their claims—when given cognitively demanding tasks from their teachers (Matsumura, Correneti, & Wang, 2015), when taught to use feedback to make their writing more persuasive (Philippakos & MacArthur, 2016), and when they participate in classroom discussions to support their argumentative writing (VanDerHeide, 2017). In the present study, cognitively stimulating VTS discussions, which incorporated peer and teacher feedback, appeared to push students beyond surface-level engagement and boosted their argument-writing skills.

Anecdotally, participants and staff reported that the STELLAR Project had broader impacts, too, altering the culture of rural schools by bringing about a level of civil discourse among teachers and students that was not previously evident. This suggests that effective

argumentation skills can generalize across academic classes and appear even in school hallways. In the context of an argument, however, this conclusion is a claim in need of further evidence in terms of theoretical and empirical support.

Limitations and Directions for Future Research

A key limitation of this work is the quasi-experimental research design without a randomized control group. Quasi-experimental designs are appropriate for small pilot studies like this, which aim to gather evidence of feasibility prior to conducting a randomized controlled trial (Kraemer et al., 2006), but they do not permit the highest degree of certainty in results. Because the intervention group teachers received the STELLAR PD and the comparison group teachers did not, it is reasonable to expect the intervention group teachers to outperform the comparison group teachers on the study's outcome measures related to the PD.

The study was limited by a relatively small sample size and geographic area, which are concerns for broad generalizability. Because data on teacher characteristics were not collected, it is not known whether teacher attributes such as teacher enthusiasm or engagement influenced outcomes. Another important limitation is that the study did not have 100% teacher compliance on a key measure of classroom implementation fidelity—eLogs. Although most intervention group teachers (70% [14/20]) submitted eLogs documenting their VTS lessons, and the eLog data provided evidence of overall intervention fidelity, a higher compliance rate would strengthen confidence in results.

The student-level results are encouraging and informative, but the lack of a student control group is an important limitation because it is impossible to compare student gains in the intervention group with student achievement under regular classroom instruction. The wide range of grade levels and subject areas in the study can be seen as both a strength and a

limitation. The study demonstrates that it is feasible to implement VTS across a broad K-12 spectrum. The broad sample also presents assessment challenges that must be considered when interpreting results. The high intraclass correlation coefficients at pretest (.34) and the low coefficients for gains in student writing scores (.04), however, suggest that students began with very different scores across grade levels but their gains over time were fairly consistent.

Taken together, the methodological limitations of this study present limitations to internal validity, and caution should be taken in interpreting and generalizing the studies' findings, but this small pilot study also had several strengths. The development and evaluation of the PD intervention at multiple, high-need rural schools under real-world conditions allows developers to consider implementation strategies and adjust for barriers (Anderson et al., 2019). The pilot study also demonstrates the feasibility of the STELLAR PD and its evaluation in the natural education settings of small, rural communities. The large effect sizes lend weight to the gains among teachers who received the PD vs. comparison teachers who did not, but the wide confidence bounds tempers enthusiasm for some results.

To better understand the benefits of teacher PD for enhancing students' visual literacy and argument writing across the curriculum, future research should test this intervention with greater rigor, perhaps incorporating more core visual arts standards into the intervention, a larger and more diverse sample, random assignment to intervention condition, more rigorous measures of implementation fidelity (such as direct observation), more attention to student writing outcomes, and more details about participants.

The conceptual theory underlying the STELLAR Project implies that student academic outcomes should improve across domains, but this initial study focused primarily on teacher outcomes and secondarily on student outcomes. Future work also might investigate whether the

in-person PD activities could be effectively transferred to an online content management system, which would help rural districts and individual schools more easily access and implement the STELLAR approach for all their students without the need for externally funded PD services and activities. Additional studies could modify and evaluate the PD program to prepare teachers specifically to teach students with learning or other disabilities and in settings such as special education classrooms. The PD could also be expanded to prepare teachers to instruct older students in the CASE criteria (complete, accurate, specific, and explained) for the effective use of evidence (Wang, Matsumura, & Correnti, 2016) or other commonly used approaches.

Conclusions

Results of this pilot study support the importance of PD to help teachers support their K-12 students in developing visual literacy and argument writing skills and in meeting CCSS and National Core Arts Standards. After receiving the STELLAR PD, teachers perceived that they were prepared to lead students in discussions about visual art through VTS, thereby helping students generate high-quality text-based arguments. Primary findings showed that teachers who completed the PD improved their perceived and actual critical thinking knowledge and skills. Secondary findings suggest that students of trained teachers made significant gains in argument writing after engaging in VTS sessions. Overall, this work indicates that investments in PD are warranted to transform classroom instruction of argument writing and to promote visual literacy, and that VTS may be an especially promising approach. With high-quality PD, teachers can more effectively promote their students' critical thinking, visual literacy, argumentation, and ultimately argument writing skills.

References

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, *20*, 93-114.

Abia-Smith, L., Smolkowski, K., Strycker, L. A., Anderson, L., & Marconi, P. (2020). *Teaching argumentation through art: Insights for museums from the STELLAR Project*. Jordan Schnitzer Museum of Art, University of Oregon.

Anderson, L., Marconi, P., Smolkowski, K, & Strycker, L. A. (2019). *Visual Teaching Strategies: A tool for teaching critical thinking and academic writing*. College of Education, University of Oregon.

Beck, S.W. (2006). Subjectivity and intersubjectivity in the teaching and learning of writing. *Research in the Teaching of English*, *40*(4), 413–460.

Cappello, M., & Walker, N. T. (2016). Visual Thinking Strategies: Teachers' reflections on closely reading complex visual texts within the disciplines. *Reading Teacher*, *70*(3), 317–325.

Cook, J., Ellerton, P., & Kinkead, D. (2018). Deconstructing climate misinformation to identify reasoning errors. *Environmental Research Letters*, *13*(024018), 1–7. https://doi.org/10.1088/ 1748-9326/aaa49f

Dash, S., de Kramer, R. M., O'Dwyer, L. M., Masters, J. & Russell, M. (2012). Impact of online professional development on teacher quality and student achievement in fifth grade mathematics. *Journal of Research on Technology in Education. 45*(1), 1–26.

De Kramer, R. M., Masters, J., O'Dwyer, L. M., Dash, S., & Russell, M. (2012). Relationship of online teacher professional development to seventh-grade teachers' and students'

knowledge and practices in English language arts. *Teacher Educator*, *47*(3), 236–259. doi: 10.1080/08878730.2012.685795

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199. doi: 10.3102/0013189X08331140

Finley, T. (2014, February 19). Common Core in action: 10 visual literacy strategies. Retrieved from http://www.edutopia.org/blog/ccia-10-visual-literacy-strategies-todd-finley

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M. & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Gilbert, M. A. (1996). How to win an argument: Surefire strategies for getting your point across (2nd ed.). New York: John Wiley & Sons.

Graff, G., & Birkenstein, C. (2014). *They say / I say: The moves that matter in academic writing* (3rd ed.). New York, NY: W. W. Norton & Co.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.

Hambek, J. (2016, March 14). Arts programs in schools often in danger of being cut. *The Washington Times*. Retrieved from: www.washingtontimes.com/news/2016/mar/14/arts-programs-in-schools-often-in-danger-of-being-/

Hillocks, G. (2011). Teaching argument writing, Grades 6-12: Supporting claims with relevant evidence and clear reasoning. Portsmouth, NH: Heinemann.

Housen, A., & Yenawine, P. (2000). *Visual Thinking Strategies basic manual Grades 3–5.* New York: Visual Understanding in Education.

Jayanthi, M., Dimino, J., Gersten, R., Taylor, M. J., Haymond, K., Smolkowski, K., & Newman-

> Gonchar, R. (2018). The impact of teacher study groups in vocabulary on teaching

> practice, teacher knowledge, and student vocabulary knowledge: A large-scale replication

> study. *Journal of Research on Educational Effectiveness*, *11*(1), 83–108. doi: 10.1080/

> 19345747.2017.1327625

Karwowski, E., Miller, A., & Sarfaty, S. (2014). Visual Thinking Strategies: An arts-based

> intervention to teach teamwork to senior medical students. *Journal of General Internal*

> *Medicine*, *29*, S543.

Klein, P., & Rose, M. (2010). Teaching argument and explanation to prepare junior students for

> writing to learn. *Reading Research Quarterly, 45*(4), 433–461. doi: 10.1598/RRQ.45.4.4

Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavage, J. (2006). Caution regarding

> the use of pilot studies to guide power calculations for study proposals. *Archives of*

> *General Psychiatry, 63*, 484–489.

Matsumura, L. C., Correnti, R., & Wang, E. (2015). Classroom writing tasks and students'

> analytic text-based writing. *Reading Research Quarterly, 50*, 417–438. doi:

> 10.1002/rrq.110

Moorman, M., Hensel, D., Decker, K. A., & Busby, K. (2017). Learning outcomes with Visual

> Thinking Strategies in nursing education. *Nurse Education Today*, *51*, 127–129. doi:

> 10.1016/j.nedt.2016.08.020

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Los Angeles:

> Muthén & Muthén.

Nanavaty, J. (2018). Using Visual Thinking Strategies with nursing students to enhance nursing

> assessment skills: A qualitative design. *Nurse Education Today*, *62*, 39–42. doi:

> 10.1016/j.nedt.2017.12.014

National Coalition for Core Arts Standards. (2014). National Core Arts Standards. Dover: State

    Education Agency Directors of Arts Education. Retrieved from:

    www.nationalcoreartsstandards.org

National Governors Association Center for Best Practices, & Council of Chief State School

    Officers. (2010). *Common Core State Standards*. Washington, DC: Author.

National Writing Project. (n.d.). *Using Sources Tool (UST)*. Retrieved from:

    https://sites.google.com/site/nwpcollegereadywritersprogram/formative-assessment-

    tools/using-sources-tool-ust

Nelson, A. (2017). Visual Thinking Strategies from the museum to the library: Using VTS and

    art in information literacy instruction. *Art Documentation: Journal of the Art Libraries

    Society of North America*, *36*(2), 281–292.

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. doi: 10.1598/RRQ.46.3.4

Olson, C. B., Scarcella, R., & Matuchniak, T. (2015). English learners, writing, and the Common

    Core. *The Elementary School Journal, 115*(4), 570–592.

Parsad, B., & Spiegelman, M. (2012). *Arts education in public and secondary schools: 1999–

    2000 and 2009–2010 (NCES 2012–2014).* Washington, DC: National Center for Education

    Statistics, Institute of Education Sciences, U.S. Department of Education.

Philippakos, Z., & MacArthur, C. (2016). The effects of giving feedback on the persuasive

    writing of fourth and fifth-grade students. *Reading Research Quarterly. 51*(4), 419–433.

    doi: 10.1002/rrq.149

State Education Agency Directors of Arts Education. (2014). *National Core Arts Standards*.

    Dover, DE: Author.

Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). New York: Cambridge University

    Press. (Original work published 1958)

Van Breukelen, G. J. P. (2006). ANCOVA versus change from baseline: More power in

    randomized studies, more bias in nonrandomized studies. *Journal of Clinical*

    Epidemiology, *59*(9), 920–925. doi: 10.1016/j.jclinepi.2006.02.007

VanDerHeide, J. (2017). Classroom talk as writing instruction for learning to make writing

    moves in literary arguments. *Reading Research Quarterly. 53*(3), 323–344. doi:

    10.1002/rrq.196

Verlaan, W. E., Ortlieb, E., & Verlaan, S. (2014). Developing argument writing through

    evidence-based responses to student-generated questions. *Voices from the Middle, 21*(3),

    34–40.

Visual Thinking Strategies. (2020). *Research and theory*. https://vtshome.org/research/

Wang, E., Matsumura, L., & Correnti, R. (2016). Making a CASE: Improving use of text

    evidence in students' writing. *The Reading Teacher, 70*(4), 479–484. doi:

    10.1002/trtr.1524

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p values: Context, process,

    and purpose. *The American Statistician*, *70*(2), 129–133. doi: 10.1080/

    00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p < 0.05."

    The American Statistician, *73*(Suppl. 1), 1–19, doi: 10.1080/00031305.2019.1583913

What Works Clearinghouse. (2017). *Procedures handbook* (Version 4.0). Washington DC: U.S.

    Department of Education, Institute of Education Sciences. Retrieved from the Institute of

    Education Sciences, National Center for Education Evaluation, WWC:

    http://ies.ed.gov/ncee/wwc/

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.

Yenawine, P. (2013). *Visual Thinking Strategies: Using art to deepen learning across school disciplines*. Cambridge, MA: Harvard Education Press.

Table 1

Profiles of Participating School Districts (2017-18)

	Comparison Districts		Intervention District		
	A	B	C	D	E
Total Enrollment	4,188	1,370	229	978	2,704
% White Teachers	95	92	100	88	91
% White Students	80	74	83	85	82
% English Learners	<5	5	<5	<5	6
% with Disabilities	16	16	17	16	16
% Free or Reduced Price Lunch	75	65	10	35	45
% Meet English Language Arts Expectations (Grade 3)	46	39	56	43	45
% On-time High School Graduation	72	85	85	88	76
Number of Teachers Providing Data	30	4	2	14	4

Note. Source of district profiles: [Oregon Department of Education At-A-Glance School and District Profiles: https://www.ode.state.or.us/data/reportcard/reports.aspx]

Table 2

Score Reliabilities (α) and Descriptive Statistics for the Four Scales from the STELLAR Teacher Survey of Critical Thinking and Argument Writing at Pretest and Posttest

Measure	Period	α	Mean (SD)	Min	Median	Max
Knowledge of Critical Thinking Concepts (CTC)						
Comparison Teachers	Pretest	.93	3.79 (0.82)	2.3	3.9	5.0
	Posttest	.88	3.65 (0.69)	1.7	3.9	4.7
Intervention Teachers	Pretest	.94	3.67 (0.83)	2.0	3.7	5.0
	Posttest	.84	4.32 (0.44)	3.6	4.3	5.0
Ability to Teach Critical Thinking Concepts (TCT)						
Comparison Teachers	Pretest	.95	3.07 (1.02)	1.0	3.3	4.8
	Posttest	.94	3.27 (0.92)	1.0	3.7	4.2
Intervention Teachers	Pretest	.95	3.06 (0.96)	1.7	2.8	5.0
	Posttest	.88	3.78 (0.60)	2.7	3.8	4.8
Knowledge of Concepts Related to Argumentative Writing (AC)						
Comparison Teachers	Pretest	.94	3.33 (1.07)	1.0	3.6	5.0
	Posttest	.93	3.40 (0.90)	1.3	3.6	4.9
Intervention Teachers	Pretest	.97	3.04 (1.18)	1.0	2.9	4.9
	Posttest	.91	3.92 (0.65)	3.0	3.9	5.0
Ability to Teach Argument Writing (TAW)						
Comparison Teachers	Pretest	.96	2.95 (0.88)	1.3	3.0	5.0
	Posttest	.96	3.40 (0.83)	1.5	3.8	4.3
Intervention Teachers	Pretest	.98	2.87 (0.99)	1.4	2.7	4.9
	Posttest	.94	3.65 (0.72)	2.4	3.4	5.0

Note. For differences between comparison and intervention teachers, all $p > .05$, and pretest differences favored the comparison condition.

Table 3

Gains on the Student Writing Assessment by Grade Level Scored with the Using Sources Tool

Measure	Gain		Standardized Gain[1]	
Presenting a claim	0.53		0.68	
Kindergarten		0.91		1.12
Grade 2		0.77		0.97
Grade 4		0.43		0.58
Grade 5		0.75		1.44
Grade 7		0.29		0.88
Grade 8		0.37		0.52
Grade 9		0.43		0.72
Grade 11		0.23		0.38
Including source material with multiple perspectives	0.57		0.83	
Kindergarten		0.36		0.67
Grade 2		0.64		1.30
Grade 4		0.46		0.85
Grade 5		1.25		1.87
Grade 7		0.29		0.48
Grade 8		0.46		0.60
Grade 9		0.83		1.52
Grade 11		0.38		0.60
Connecting source material to the claim	0.69		0.79	
Kindergarten		0.55		0.82
Grade 2		0.82		1.98
Grade 4		0.63		0.94
Grade 5		1.17		1.30
Grade 7		0.82		1.09
Grade 8		0.48		0.55
Grade 9		0.57		0.61
Grade 11		0.62		0.90
Organization contributing to argument development	0.75		0.80	
Kindergarten		0.82		1.19
Grade 2		0.93		1.41
Grade 4		0.66		0.80
Grade 5		1.17		1.48
Grade 7		1.06		1.24
Grade 8		0.51		0.58
Grade 9		0.70		0.68
Grade 11		0.46		0.61

Measure	Gain		Standardized Gain[1]	
Language contributing to argument development	0.56		0.68	
Kindergarten		0.82		1.62
Grade 2		0.57		0.88
Grade 4		0.51		0.76
Grade 5		1.00		1.25
Grade 7		0.47		0.67
Grade 8		0.49		0.58
Grade 9		0.43		0.53
Grade 11		0.69		1.24
Writing mean (average of five subscales)	0.62		0.87	
Kindergarten		0.69		1.57
Grade 2		0.75		1.47
Grade 4		0.54		0.95
Grade 5		1.07		1.60
Grade 7		0.59		1.13
Grade 8		0.46		0.64
Grade 9		0.59		0.79
Grade 11		0.48		0.84
Word count	23.9		0.42	
Kindergarten		23.9		2.05
Grade 2		33.6		1.54
Grade 4		20.7		1.09
Grade 5		68.3		2.00
Grade 7		7.2		0.19
Grade 8		10.9		0.31
Grade 9		23.9		0.22
Grade 11		18.9		0.85

[1]The standardized gain was calculated as the gain divided by the pretest standard deviation.

Visual Thinking Strategies and Argument Writing

Appendix A

National Writing Project Assessment Tool and Student Writing Sample Scoring Sheet

Student writing samples were scored with the National Writing Project's (n.d.) *Using Sources Tool*. For all writing samples raters completed a scoring sheet, shown in Figure A1, on the six dimensions listed next with the definitions for each point along the scale given below.

Using Sources Tool Dimensions and Scale

A. Does the writing present a claim?
1. The writing presents a claim that is nuanced, debatable, and defensible. (Nuanced = multiple layers of detail, pattern, or meaning.)
2. The writing presents a claim that is debatable and defensible.
3. The writing presents a summary statement about source material, but that statement is not debatable.
4. The writing does not present a claim.

B. Does the writing include source material that represents multiple perspectives?
1. Not present. (The writing does not use source material.)
2. Developing. (The writing summarizes source material without connecting it to the claim.)
3. Competently. (The writing uses source material to support the claim.)
4. Effective. (The writing skillfully integrates source material to fully support the paper's claim.)

C. Does the writing comment on source material in ways that connect the source material to the claim?
1. Not present. (The writing contains no commentary on source material.)
2. Developing. (The commentary summarizes the source material and/or offers little analysis to connect to the claim.)
3. Competently. (The commentary sometimes includes analysis that shows how the source material connects to and supports the claim.)
4. Effective. (The commentary consistently includes analysis that clearly shows how the source material connects to support the claim.)

D. Does the organization of the writing contribute to the overall development of argument?
1. Not present. (The writing is disjointed and the logic and reasoning are unclear.)
2. Developing. (The writing is somewhat organized and the reader can somewhat follow the logic and reasoning.)
3. Competently. (The writing is somewhat organized and the reader can follow the logic and reasoning.)
4. Effective. (The writing is thoughtfully organized and the ideas flow smoothly. The reader can easily follow the logic and line of reasoning.)

E. Does the language of the writing contribute to the overall development of argument?
1. Not present. (The writing contains limited vocabulary and words that are frequently used incorrectly.)
2. Developing. (The writing contains words that are occasionally clear and precise, may be vague and used incorrectly.)
3. Competently. (The writing contains words that are clear and precise, appropriate and varied. The writing may include imagery.)
4. Effective. (The writing contains words that are powerful, vivid, and varied. The writing also contains vivid verbs, precise nouns, and modifiers that add depth and specificity to the writing.)

F. Word Count

Scale Point Definitions

Effectively = The writing makes the move (i.e., distinguishing student's and a source's ideas, selecting evidence to the support the claim, connecting evidence to the claim, or characterizing credibility of a source) in a way that contributes to the overall development of the paper. When present, the move is purposeful, logical, and consistent.

Competently = The writing makes the move (i.e., distinguishing student's and a source's ideas, selecting evidence to the support the claim, connecting evidence to the claim, or characterizing credibility of a source). The move is generally controlled and satisfactory. There may be occasional lapses, if a move is attempted more than once.

Developing = The writing attempts the move (i.e., distinguishing student's and source's ideas, selecting evidence to the support the claim, connecting evidence to the claim, or characterizing credibility of a source), but may do so in a limited or underdeveloped way. If a move is attempted more than once, its use may be very uneven.

Not present = There is no evidence of a particular move in a paper (i.e., distinguishing student's and source's ideas, selecting evidence to the support the claim, connecting evidence to the claim, or characterizing credibility of a source). Alternatively, the writing cannot be evaluated for a particular move because it summarizes or copies without attribution or may be too brief to evaluate.

Figure A1

STELLAR Student Writing Sample Scoring Sheet

Paper #	Reader #	School #	Student #	Grade Level	Present a claim?	Source material that represents multiple perspectives?	Connect the source material to the claim?	Organization contributes to overall dev. of argument?	Language contributes to the overall dev. of argument?	Word Count
1										
2										
⋮					⋮					⋮