

# Comprehension and Writing Strategy Training Improves Performance on Content-Specific Source-Based Writing Tasks

Jennifer L. Weston-Sementelli<sup>1</sup> · Laura K. Allen<sup>1</sup> ·  
Danielle S. McNamara<sup>1</sup>

Published online: 9 December 2016

© International Artificial Intelligence in Education Society 2016

**Abstract** Source-based essays are evaluated both on the quality of the writing and the content appropriate interpretation and use of source material. Hence, composing a high-quality source-based essay (an essay written based on source material) relies on skills related to both reading (the sources) and writing (the essay) skills. As such, source-based writing must involve language comprehension and production processes. The purpose of the current study is to examine the impact of reading, writing, and blended (i.e., reading and writing) strategy training on students' performance on a content-specific source-based essay writing task. In contrast to general source-based writing tasks, content-specific source-based writing tasks are tasks wherein writers are provided the source material on which to base their essays. Undergraduate students ( $n = 175$ ) were provided with strategy instruction and practice in the context of two intelligent tutoring systems, *Writing Pal* and *Interactive Strategy Training for Active Reading and Thinking* (iSTART). Results indicated that participants in the blended strategy training condition produced higher quality source-based essays than participants in the reading comprehension-only, writing-only, or control condition, with no differences observed between the latter three conditions. Further, the benefits of this blended strategy instruction remained significant regardless of prior reading and writing skills, or time on task.

**Keywords** Source-based writing · Content area writing · Reading · Writing · Intelligent tutoring systems

---

✉ Danielle S. McNamara  
dsmcnamara1@gmail.com

<sup>1</sup> Arizona State University, Tempe, AZ, USA

## Introduction

Writing takes many forms. In our daily lives, we write notes, email messages, tweets, and blogs. As professionals and academics, we write reports, chapters, and journal articles (such as this one). As students, we are required to write essays that demonstrate our writing proficiencies. These assignments take several forms. Students may be asked to write about what they did over the summer, discuss the consequences of a particular historical event, or express their opinions on political, scientific, or pop culture issues. To assess writing skills, educators and researchers often use persuasive essays that prompt students to discuss their opinions on various topics, primarily because this task generally does not require students to utilize source material or have significant prior knowledge of a particular domain. By contrast, source-based essays ask students to read material and answer integrative questions about a particular topic. These essays are generally assigned in content area courses such as science, social studies, history, and literature, wherein the student needs to read and integrate information across multiple texts to answer one or more questions. These essays are also increasingly being used to assess students' writing skills.

Writing in a discipline involves describing, summarizing, and integrating information related to that discipline (e.g., science, history) and presenting new ideas related to those concepts. Developing writers learn to engage in disciplinary writing by learning to paraphrase and summarize ideas, and to integrate these ideas to address one or more questions. Additionally, these writers must gain knowledge of the disciplinary content through various sources, such as listening to a lecture or reading a text. One type of essay commonly used in educational settings to provide instruction toward this objective, and to assess students' ability to engage in disciplinary writing is the source-based essay. The term "source-based writing" can refer to a wide variety of written tasks, including summaries, reaction papers, syntheses, lab reports, constructed responses, argumentative papers, research papers, and essay exam questions. Source-based writing differs from other forms of writing (e.g., persuasive or narrative writing) because it requires the writer to synthesize information from texts in response to a prompt or goal (Braine 1995; Eblan 1983).

Students' success on these source-based writing tasks relies on their understanding of the content<sup>1</sup> in a particular domain, as well as their ability to accurately convey this knowledge. To be considered high quality, source-based essays must show mastery of the conventions of writing, an accurate understanding of the source material, and utilize the material appropriately, presenting a synthesis of the material in response to the question. Consequently, performance on these essays serves as an important indicator of students' knowledge and skills<sup>2</sup> in academic settings. Little empirical research, however, has been conducted to understand the cognitive processes<sup>3</sup> necessary to produce these source-based essays (beyond those required for reading and writing independently), nor the pedagogical techniques that most effectively improve students' ability<sup>4</sup> to compose source-based essays. The latter is the focus of this research study.

<sup>1</sup> Content here refers to both prior knowledge and information extracted from sources.

<sup>2</sup> Skills refers to the proficiencies that one develops to complete a task that come from training, experience, or practice.

<sup>3</sup> Cognitive Processes refers to higher mental processes, such as perception, memory, language, problem solving, and abstract thinking involved in completing a task.

<sup>4</sup> Ability refers to the possession of the means to complete a task.

Specifically, our aim is to examine the impact of different types of strategy instruction (i.e., reading comprehension and writing) on source-based essay quality.

### Source-Based Essays

The present study focuses on source-based essays requiring the use and synthesis of multiple sources where the response is evaluated both on the quality of the writing and the content included. Specifically, this study utilizes *content-specific* source-based essays, where writers are provided the sources on which to base their essays. To succeed on these writing tasks, writers need to not only have strong writing skills, but also to be able to read and understand the sources provided, such that the content of the essays is accurate. As such, source-based essay writing tasks rely on proficiency in both writing and reading.

Unfortunately, national and international assessments suggest that many students struggle in the domains of both writing and reading. For example, the 2011 National Assessment of Educational Progress (NAEP) report indicated that 21 % of high school seniors failed to meet basic proficiency standards in writing and only 26 % met the standards for proficiency. Likewise, the 2013 NAEP report showed that a majority (64 %) of high school seniors scored at or below basic proficiency in reading. These trends are far worse for minorities and English as a Second Language (ESL) students.

Considering high school students' overall lack of literacy proficiency, it is understandable that they would particularly struggle with source-based essay writing. Unfortunately, little empirical research has been conducted on the processes involved in writing source-based essays, nor on how to most efficiently improve these skills. Researchers across multiple disciplines have examined students' ability to comprehend multiple documents, integrate knowledge, and evaluate source materials using different forms of source-based writing (e.g., Anmarkrud et al. 2014; Gerard et al. 2016; Linn et al. 2003; Wiley et al. 2009; Wiley et al. 2014). Importantly, however, these studies do not focus on the processes involved in producing source-based essays; rather, the essays are taken as a means of assessing students' text comprehension (e.g., Britt and Aglinskas 2002; Linn et al. 2003; Rouet et al. 1996; Wiley et al. 2014; Wineburg 1991).

Few researchers have directly investigated source-based essay writing as an academic task while focusing on both the content and the *compositional quality* of what is written. Although prior research has examined the relations between the quality of persuasive essays and their linguistic features (e.g., syntactic complexity, language sophistication, cohesion, concreteness, pronoun usage; Crossley et al. 2011; McNamara 2013; McNamara et al. 2010), comparable studies of source-based writing are not known to us. Those who have examined source-based writing (even short-answer questions) have generally focused on the proximity of the content to the sources, quality of argumentation, and on the selection of source materials (Anmarkrud et al. 2014; Britt and Aglinskas 2002; Foltz et al. 1996; Rouet et al. 1996; Wiley et al. 2009). While these aspects of source-based writing are clearly important, prior studies have not focussed on compositional quality and have not described the criteria for quality (Britt and Aglinskas 2002; Wiley et al. 2009; Wiley et al. 2014).

In addition, the majority of prior research has focused on the processes involved in the understanding, integration, or evaluation of sources, rather than on the processes

involved in generating the essays (e.g., Anmarkrud et al. 2014; Stadler and Bromme 2007; Wiley et al. 2014). Finally, much of the prior literature on source-based writing has targeted the building of a specific understanding to which there is a “correct” answer (e.g., Gerard et al. 2016; Stadler and Bromme 2007; Wiley et al. 2009); by contrast, source-based essay writing tasks do not generally have a single correct answer or single correct interpretation of the source material.

Importantly, only a small number of studies have examined the effects of instruction or training on students’ ability to utilize source material (Britt and Aglinskas 2002; Foltz et al. 1996). Those that have done so have focused on the effects of training on “sourcing” (i.e., students’ ability to evaluate source material and select relevant information in response to questions; e.g., Britt and Aglinskas 2002; Rouet et al. 1996). However, they have not targeted problems that students may experience regarding their basic comprehension of the sources or the processes involved in the production of well-written essays or short-answer questions.

## Reading and Writing

Composing a high-quality source-based essay (based on both content and compositional quality) relies on both reading (the sources) and writing (the essay) skills, and therefore must involve language comprehension and production processes. Already, there is inherent overlap between reading and writing; they are, for example, the primary forms of text-based communication. When writers produce text, they use their knowledge of the world to communicate (or construct) meaning for a particular audience; similarly, readers construct meaning by interpreting texts based on their own prior knowledge and goals (Spivey 1990). Indeed, educators and researchers commonly assume that reading and writing rely on common knowledge and processes (Fitzgerald and Shanahan 2000; Tierney and Shanahan 1991), and many studies have found correlations between students’ performance on reading and writing tasks (e.g., Loban 1967).

Though correlations between reading and writing measures typically never exceed .50 (Tierney and Shanahan 1991), one important question regards the source of the overlap – namely, which processes and knowledge are common to both reading and writing? Most studies investigating this question find strong overlap in the lower level processes such as phonemic awareness and vocabulary knowledge, but a weaker overlap in the higher levels such as discourse knowledge, strategy knowledge, and inferencing abilities (Allen et al. 2014b; Allen, Perret, & McNamara, 2016; Berninger et al. 2002, Juel et al. 1986).

At the same time, providing instruction at these higher levels improves comprehension and writing. Providing students with instruction and practice in using reading strategies and generating inferences improves students’ comprehension (Brown 1982; Palincsar and Brown 1984; McNamara 2004). Providing students with writing strategy instruction improves students’ holistic scores on writing tasks (Allen et al. 2014a; Graham and Perin 2007), which take into account ideation, organization, vocabulary, and sentence structure (Diederich 1966).

Nonetheless, while a good deal of evidence suggests that reading and writing strategy training improve reading and writing performance, respectively, there is no evidence of their effectiveness on source-based essay tasks. What type of strategy

instruction is most beneficial for students' success on source-based writing tasks? And, do the effects of instruction depend on students' prior reading or writing skills?

To address these questions, this study examines the differential effects of providing students with instruction and practice on reading comprehension strategies, writing strategies, or a combination of both comprehension and writing strategies. We also examine the degree to which the benefits of instruction depend on students' prior abilities in reading and writing. If students struggle with basic writing strategies, the quality of their source-based essays may improve with writing strategy training more so than with comprehension strategy training. Similarly, if students struggle with comprehension skills, the quality of their source-based essays may be enhanced with comprehension strategy training more so than with writing strategy training. Still, given the moderate correlation between reading and writing skills, as well as their mutual contributions to source-based writing, performance might be expected to benefit from a combination of writing and reading strategy training regardless of individual differences in reading or writing skill. To investigate these potential outcomes, we turn to two intelligent tutoring systems (ITSs), iSTART and the Writing Pal, which afford the delivery of automated, adaptive instruction and practice to students on reading comprehension and writing strategies.

## **iSTART**

iSTART is an ITS designed to enhance students' reading comprehension skills through instruction on self-explanation and reading comprehension strategies, such as comprehension monitoring, paraphrasing, prediction, bridging, and elaboration (McNamara 2004). Self-explanation is the process of explaining the meaning of text to oneself, particularly by grounding text information in prior knowledge. The relationship between self-explanation and comprehension strategy instruction is symbiotic in the sense that self-explanation externalizes the use of the strategies for the students, and at the same time prompts a focus on causal relationships within the text, which enhances students' bridging and elaborative inferences. Self-explanation is also akin to other techniques (e.g., summarization, question answering) that encourage readers to write about what they read (Hebert et al. 2013; Newell 2007). Prompting readers to write about what they have read impacts understanding by fostering explicit knowledge and the construction of relationships between ideas, allowing readers to compare what they have written to other sources. Overall, the purpose of the strategies taught in iSTART is to improve students' understanding of text meaning by encouraging them to establish connections between the concepts in a text, as well as with information outside the text (McNamara et al. 2007b).

iSTART and its non-automated predecessor, Self-Explanation Reading Training (SERT) have been shown to effectively improve strategy use, reading comprehension, and course performance for a range of students from middle school to college (Jackson and McNamara 2013; Magliano et al. 2005; McNamara 2004, 2015; McNamara et al. 2004; McNamara et al. 2006; McNamara et al. 2007a, c; Snow et al. 2016). iSTART includes instructional videos, Coached Practice, and a suite of both generative and identification games. The instructional videos use animated agents to provide the initial instruction in self-explanation and the comprehension strategies. Within iSTART, there are a total of five strategies: comprehension monitoring, predicting, paraphrasing,

elaborating, and bridging. Each strategy is explained and demonstrated within a single five minute video and at the end of each video students take a short multiple-choice quiz referred to as a checkpoint that is designed to assess their understanding of the recently learned strategy. Checkpoint completion is used to track progress through the system and some features will not unlock if a checkpoint has not been completed. In Coached Practice,<sup>5</sup> students are guided through a text and prompted to self-explain target sentences. Students receive scaffolded feedback from an animated agent after each self-explanation. After students complete the training phase of iSTART, they are transitioned to the practice phase, which contains an interactive game-based interface. This interface provides students with the opportunity to practice using the self-explanation strategies they learned in the training phase. Within the practice activities, students can read and self-explain new texts, play mini-games, personalize the background color of the interface or customize their avatar, and monitor their own performance within the system.

Within the game-based interface, there are two types of practice in which students can engage: generative and identification. In generative practice games, students read scientific texts and type self-explanations in response to several target sentences. In identification practice games, students read science texts and self-explanations that were written by other students, and then identify which of the five strategies were used to generate that self-explanation. There are three generative practice environments within iSTART: Coached Practice, Showdown, and Map Conquest. The game versions of generative practice (Showdown and Map Conquest) are designed to engage students' interest while they practice generating self-explanations. For example, in Map Conquest students are asked to generate a self-explanation for numerous target sentences while collecting flags they can use to conquer a map. Students can earn flags in this game by generating high quality self-explanations.

Students' generated self-explanations are scored using a computational algorithm, using natural language processing, which assigns a score between 0 and 3 to each self-explanation. This algorithm uses Latent Semantic Analysis (LSA; Landauer et al. 2007) and word-based measures to assess self-explanation quality (Jackson et al. 2010b; McNamara et al. 2007c). Higher scores are assigned to self-explanations that use key words and include language related to the text content (both the target sentence and previously read sentences), whereas lower scores are assigned to unrelated or short responses. The scoring algorithm thus intends to reflect how well students have established relevant connections between the target sentence and prior text material and prior knowledge.

## The Writing Pal

The Writing Pal is an ITS designed to provide explicit instruction on writing strategies and to provide students with opportunities to practice writing and receive feedback. The system was specifically designed to target the writing of persuasive essays, similar to those found on many standardized tests (Roscoe and McNamara 2013). In these tasks, students are provided with a prompt that describes a question that can be debated using

---

<sup>5</sup> The first instance of coached practice can be considered part of the training and is generally completed prior to full access to the game-based practice environment.

evidence from experience or common world knowledge. For example, students might be provided with a prompt that introduces the notions of cooperation versus competition in achieving goals, and asked to write an essay on whether people achieve more success through cooperation or by competition. These essays differ from source-based essays in that they do not require (or support) the usage of source materials.

The Writing Pal provides instruction and practice on basic writing strategies that have been found to improve students' performance on persuasive essays. Instruction includes a series of nine strategy instruction modules, covering freewriting, planning, introduction building, body building, conclusion building, cohesion building, and revision. A large body of educational research supports the importance of these strategies to writing (Cameron and Moshenko 1996; Faigley and Witte 1981; Flower and Hayes 1980; Graesser et al. 2004; Graham and Harris 2000; Henry and Roseberry 1997; McCarthy et al. 2008; Zimmerman and Risemberg 1997). Students are taught explicit strategies for generating and organizing their ideas, drafting persuasive essays with a clear rhetorical structure, and revising their essays to express ideas in a more sophisticated and cohesive manner. Strategy instruction is provided via 5–10 min lesson videos presented by animated characters; at the end of each video, students take a short quiz or checkpoint that is designed to assess their understanding of the recently learned strategy. Checkpoint completion is used to track progress through the system and some features will not unlock if a checkpoint has not been completed. Students are offered two related modes for practice. The strategy lessons are associated with over a dozen unique game-based practice activities that enable students to practice specific strategies. In addition, students practice writing complete essays using the automated writing evaluation (AWE) component, which provides automated summative and formative feedback to guide their overall strategy use and learning (McNamara et al. 2015a). The essay AWE component for Writing Pal allows learners to draft essays, receive targeted feedback, and revise essays (receiving more feedback at the second submission).

The suite of games available in the Writing Pal includes games targeting both the identification of strategy usage, and the generation text aligned with specific strategies. Identification games target skills such as planning, attention grabbing strategies, and cohesion. Generative games include games targeting the writing of topic and evidence sentences and the improvement of essays through revision. Feedback is provided in generative games using algorithms based on the use of natural language processing.<sup>6</sup>

The Writing Pal covers the entire writing process from idea generation through revision; however, it is designed to be modular, allowing educators and researchers to utilize only the modules they deem necessary. This feature allows educators to target skills they believe their students may be lacking. Furthermore, as large class sizes have limited the ability of teachers to provide frequent writing practice with feedback to students (National Commission on Writing 2003), the utility of automated writing tutors such as Writing Pal has increased. The present study capitalizes on the modular aspect of the Writing-Pal system to provide instruction and practice to the students most relevant to source-based writing.

The impact of the Writing-Pal instruction on writers has been positive. For instance, Roscoe et al. (2014) reported that students who utilized the Writing-Pal were more

<sup>6</sup> For more information on the linguistic features used to provide feedback in the games see Roscoe et al. 2013

likely to learn new strategies compared to a writing-only control group. Further, Writing Pal training has been linked to increases in students' holistic essay scores on persuasive SAT-style writing tasks over time (Allen et al. 2014a, 2015; Crossley et al. 2013) scored using an algorithm based on expert scores of SAT-style essays using the SAT writing rubric. Furthermore, the game based practice available in the Writing Pal has been shown to be engaging for students, which is a key factor in persistence (Roscoe et al. 2013).

## Method

The objective of this study is to examine the differential impacts of providing students with instruction and practice on strategies to improve reading comprehension, writing, or both comprehension and writing, compared to a control condition. The study included two sessions. The first session comprised a pretest assessing initial reading and writing ability, as well as strategy training for experimental conditions. During the second session, participants completed a timed source-based writing task. Sessions occurred between one and three days apart to accommodate the scheduling needs of participants.

## Participants

Undergraduate psychology students from Arizona State University participated in this study for credit in their Psychology 101 course and were randomly assigned to one of the four conditions: no instruction control group, iSTART only, Writing Pal only, or combined instruction (both iSTART and Writing Pal training). Of the 232 participants for whom complete data was collected,<sup>7</sup> this study examines performance for the 175 participants ( $n_{Control} = 48$ ,  $n_{iSTART} = 41$ ,  $n_{Writing Pal} = 41$ ,  $n_{Combined} = 45$ ) who identified English as their first language. Participants ranged in age from 17 to 43 with a mean age of 19.6 years (median age = 19; SD = 3.4). Half of participants were freshmen (50 %) and 57 % were male. Participants reported a number of ethnic backgrounds with the majority being Caucasian (66 %), Hispanic (16 %), or Asian (7 %) decent.

## Procedure

During Session 1, students first completed a pretest, which was comprised of demographic, motivation, and self-efficacy measures.<sup>8</sup> Participants then completed a timed (25-min) SAT-style essay followed by the Gates MacGinitie Reading Test. The trajectory of each student following these initial assessments varied based on their condition (see Table 1). All conditions were designed to take approximately 3 h, however,

<sup>7</sup> The present study examines a subset of participants that includes only those who identified English as their first language, completed both sessions with complete data. A total of 261 participants began the study; 10 participants did not complete both sessions, 18 participants did not have complete data, and 58 reported English as their second language

<sup>8</sup> Motivation and self-efficacy measures are not discussed in the present study because preliminary analyses suggested no impact of motivation and self-efficacy on source-based essay scores.



**Table 1** Flow of source-based essay study

Condition		Sections counterbalanced		
	Writing Pal only	Combined (Writing-Pal section)	Combined (iSTART section)	Control
iSTART only				
Session 1 (3 h)				
Demographics				
Motivation and self-efficacy measures				
25-min SAT-style essay				
Gates MacGinitie Reading Test				
Overview video	Positions, Arguments, and Evidence video Play Planning Passage twice	Positions, Arguments, and Evidence video Play Planning Passage	Overview video	Prior knowledge test
Comprehension Monitoring video	Play Planning Passage twice	Play Planning Passage	Comprehension Monitoring video	Gates MacGinitie Vocabulary Test
Prediction video	Thesis Statements video	Thesis Statements video	Prediction video	Component processes test
Paraphrasing video	Argument Preview video	Body Building overview video	Paraphrasing video	Symmetry Span
Elaboration video	Body Building overview video	Evidence Sentences video	Elaboration video	Operation Span
Bridging video	Topic Sentences video	Strengthening your Evidence video	Bridging video	Flanker Attention Control Task
Summary video	Evidence Sentences video	Play RoBoCo for 10 min	Summary video	
Demonstration video	Strengthening your Evidence video	Conclusion Building Overview video	Demonstration video	
45 min of Coached Practice <sup>a</sup>	Play RoBoCo for 20 min	Summarize the Essay video	1 text in Coached Practice	
45 min of practice (users choice) <sup>a</sup>	Conclusion Building Overview video	Play Lockdown for 10 min	user choice of practice in iSTART	
	Summarize the Essay video			
	Play Lockdown for 20 min			

**Table 1** (continued)

Condition	
Sections counterbalanced	
iSTART only	Writing Pal only
iSTART only End session 1 Motivation assessment Introduction to source-based writing Self-efficacy measure 40-min source-based writing task	Combined (Writing-Pal section) Combined (iSTART section) Control

<sup>a</sup> Because some students take longer than anticipated to complete the previous tasks the remaining time was split between coached practice and user choice practice, with the participants completing at least one text in coached practice

completion time for session one varied ranging from roughly 1.5 h to 3.5 h. The actual completion time varied for a number of reasons including condition. In general, the control condition took less time to complete than the other conditions, as the tasks were entirely user-paced. Time to complete training in the tutoring systems also varied with some students attempting to game the system to finish earlier. As such, the impact of time spent in interacting with the systems on source-based writing was assessed.

Following this initial training session, participants returned between 1 and 3 days<sup>9</sup> later to complete session two. In this session, participants completed a motivation questionnaire prior to being introduced to the source-based writing task. The essay prompt was provided and participants then completed a task-specific self-efficacy questionnaire. Following the questionnaire, participants were directed to a web page containing the prompt, sources, and a Word document to download and write in. They were shown how to access the sources and use the split screen function to allow for simultaneous viewing of their essay and the source materials. Finally, participants were given 40 min to complete the source-based essay task. Following the study, participants were thanked for their time and debriefed. Session two took between 45 and 60 min for participants to complete depending on the amount of time spent completing the motivation and self-efficacy measures.

### *Control Group*

Those in the control condition completed a prior knowledge test and the Gates MacGinitie Vocabulary Test (GMVT, MacGinitie and MacGinitie 1989) prior to completing a series of working memory and attention control tasks to control for time on task. The results from these tasks are not discussed in the present study.

### *Training Conditions*

**iSTART** For the present experiment, participants watched all seven instructional videos (i.e., five targeting the previously mentioned strategies, along with an overview and summary video), completed the corresponding checkpoints (short multiple choice knowledge checks), the demonstration video, completed a text in Coached Practice, and had access to the suite of games. iSTART was designed to aid students in reading science texts; however, the strategies taught are applicable to any kind of text. All of the games targeting the application or identification of self-explanation strategies were available to participants in this study. Because the available games target all of the self-explanation strategies, participants were given the ability to choose which games they completed.

Because iSTART instruction requires less time than the Writing Pal to complete, and all lesson videos are applicable to understanding source materials, participants in this study viewed all of the lesson videos along with the demonstration video. Participants

---

<sup>9</sup> 97 % of participants completed session two within 3 days; due to scheduling constraints the remaining 3 % (5 participants) completed their second session outside of this window.

in the iSTART only condition then split their time<sup>10</sup> between Coached Practice and free choice within the environment (access to Coached Practice and games<sup>11</sup>). Following their time interacting with Coached Practice, participants were given free-choice of games and coached practice within iSTART for their remaining session time because only two different tasks were actually available to them, practice involving the identification of self-explanation strategies and generative practice.

**The Writing Pal** The Writing Pal provides instruction and practice on strategies related to performance on SAT-style persuasive essays. Hence, not all of the lesson videos and games are directly applicable to source-based writing. However, the modular aspect of the Writing Pal affords selecting videos and games based on the nature of the task. A total of nine videos (with corresponding checkpoints) and three games from four different modules (i.e., Planning, Introduction Building, Body Building, Conclusion Building) were selected for inclusion in this study. Specifically, this study includes lessons covering the topics of: *Positions, Arguments, and Evidence* (planning), *Thesis Statements* (introductions), *Argument Previews* (introductions), *Topic Sentences* (body paragraphs), *Evidence Sentences* (body paragraphs), *Strengthening Evidence* (body paragraphs), *Conclusion Building* (conclusions), and *Summarizing* (conclusions). Two generative games (*RoBoCo* and *Lockdown*) and one identification game (*Planning Passage*) were also selected to be included in this study. Participants did not interact with the essay writing module during this study.

The strategy lessons selected to be used in this study are lessons expected to be applicable to source-based writing, and are framed in a way that makes them applicable to essentially any kind of writing. When writing a source-based essay it is crucial that the writer selects appropriate evidence from the sources. For this reason, the majority of the lessons discuss evidence in some fashion. Strong introduction and conclusion paragraphs are also critical to any successful essay; as such, lessons targeting critical parts of these paragraphs were used.

The use of the game-based practice available in the Writing Pal has been shown to enhance strategy acquisition, engagement, and motivation (Allen et al. 2014a). The range of games appropriate for the present study was limited because many of the games rely on the player having seen all of the lessons in the module. One identification and two generative games were selected for this study from three different modules, Planning, Body Building, and Conclusion Building. Planning passage is an identification game wherein players identify the appropriate arguments for a position, and the appropriate evidence to support an argument. The two generative games used in this study were *RoBoCo* and *Lockdown*; these games require the player to construct responses in natural language. In *RoBoCo*, the player builds robots by writing topic and evidence sentences given a thesis statement. In *Lockdown*, players are asked to

<sup>10</sup> Because some students took longer than anticipated to complete the previous tasks, the remaining time was split between coached practice and user choice practice, with the participants completing at least one text in coached practice.

<sup>11</sup> Participants were locked out of the non-practice components of iSTART.

write a conclusion paragraph based on an outline; a high quality conclusion paragraph serves to stop computer hackers.

### *Combined Instruction*

The combined group completed an abridged version (1 h in each system) of both the Writing Pal and iSTART training and the order of presentation of the Writing Pal and iSTART was counterbalanced. In the combined condition, participants played games for a shorter period of time and did not view the argument preview or topic sentence videos in the Writing Pal. During iSTART training, participants were required to complete one text in Coached Practice, and were given free access to the system if they had time remaining.

## **Measures**

### *Writing Proficiency*

Writing proficiency was assessed at pre-test using a 25-min SAT-style persuasive essay that participants completed prior to training. The essay used on the SAT<sup>12</sup> is designed to measure a students' ability to take a position on a prompt and support it in writing. Holistic essay scores are based on quality, not length ([collegereadiness.collegeboard.org/sat-essay-scoring-before-march-2016](http://collegereadiness.collegeboard.org/sat-essay-scoring-before-march-2016)) with a focus on features such as the use of appropriate examples and evidence, organization, coherence, use of language and vocabulary, and the absence of errors in grammar, usage, and mechanics. The essays were completed in Qualtrics instead of The Writing Pal as not all participants interacted with The Writing Pal. Unlike essays completed in The Writing Pal, participants in the present study did not receive feedback on their pre-test essays. The essay was automatically submitted after 25 min (the interface provided a visible timer) and participants were not able to submit the essay early. The prompts utilized to assess prior writing ability are SAT style persuasive essay prompts (obtained from [onlinemathlearning.com](http://onlinemathlearning.com)). Two prompts were utilized in the present study to control for potential prompt effects. Half of the participants in each condition were assigned to each pre-test prompt.<sup>13</sup> Participants were instructed: *You will now have 25 min to write an essay on the prompt below. The essay gives you an opportunity to show how effectively you can develop and express ideas. You should, therefore, take care to develop your point of view, present your ideas logically and clearly, and use language precisely. Think carefully about the issue presented in the following excerpt and the assignment below. [Insert 1 Prompt from below] Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations.*

The two prompts used in this study are from retired SAT exams and have been minimally edited to increase clarity.

<sup>12</sup> This study utilizes the SAT-style essay and scoring guide used prior to March 2016

<sup>13</sup> Participants were prompted raise their hand prior to proceeding to the pretest essay for an experimenter to check that they had entered their ID information correctly, unfortunately 17 participants did not follow directions and completed the alternate pretest prompt.

## Images and Impressions

All around us appearances are mistaken for reality. Clever advertisements create favorable impressions but say little or nothing about the products they promote. In stores, colorful packages are often better than their contents. In the media, how certain entertainers, politicians, and other public figures appear is sometimes considered more important than their abilities. All too often, what we think we see becomes far more important than what really is.

Do images and impressions have a positive or negative effect on people

## Competition and Cooperation

While some people promote competition as the only way to achieve success, others emphasize the power of cooperation. Intense rivalry at work or play or engaging in competition involving ideas or skills may indeed drive people either to avoid failure or to achieve important victories. In a complex world, however, cooperation is much more likely to produce significant, lasting accomplishments.

Do people achieve more success by cooperation or by competition?

The essays were scored holistically using an algorithm currently utilized in the Writing Pal for college aged students. This algorithm was developed based on expert ratings of 1234 similar essays using the 6-point rating scale developed by the SAT. This rubric (see Table 2) is not tied to a specific prompt but designed to be used to score all argumentative essays on the SAT. Scores based on this scoring rubric are tied to general features of writing including sophistication of vocabulary, evidence-based reasoning, coherence, varied sentence structure, and attention to the conventions of English (SAT rubric available from: [collegereadiness.collegeboard.org/sat-essay-scoring-before-march-2016](http://collegereadiness.collegeboard.org/sat-essay-scoring-before-march-2016)). The essay scoring algorithm scores each essay on a 1 to 6 scale (similar to the SAT rating scale). Three research instruments, Coh-Metrix (Graesser et al. 2004; McNamara and Graesser 2012; McNamara et al. 2014), the Writing Analysis Tool (McNamara et al. 2013), and LIWC (Pennebaker et al. 2007) were used to assess essays on hundreds of different linguistic indices including indices of cohesion, connectives, lexical and semantic co-referentiality, causal cohesion, lexical diversity, spatiality, temporality, paragraph cohesion, vocabulary, word frequency, word information measures, n-grams, nominals, verb-related features, syntactic indices, rhetorical and semantic features, lexical features, psychological semantics, and narrativity. These linguistic indices were then correlated with expert scores for essays to determine which indices were most related to expert judgements of quality. A step-wise discriminant function analysis was used to classify essays and resulted in a hierarchical algorithm for assessing SAT-style persuasive essays. The accuracy of the AWE system utilized by The Writing Pal has been found to be equivalent to expert accuracy, with 44–55 % exact and 94–96 % adjacent accuracy (within one score point) with expert scores (McNamara et al. 2015a).

**Table 2** SAT persuasive essay scoring rubric

Score	Criteria
6	<p>An essay in this category demonstrates clear and consistent mastery, although it may have a few minor errors. A typical essay:</p> <ul style="list-style-type: none"> <li>• Effectively and insightfully develops a point of view on the issue and demonstrates outstanding critical thinking, using clearly appropriate examples, reasons and other evidence to support its position</li> <li>• Is well organized and clearly focused, demonstrating clear coherence and smooth progression of ideas</li> <li>• Exhibits skillful use of language, using a varied, accurate and apt vocabulary</li> <li>• Demonstrates meaningful variety in sentence structure</li> <li>• Is free of most errors in grammar, usage and mechanics</li> </ul>
5	<p>An essay in this category demonstrates reasonably consistent mastery, although it has occasional errors or lapses in quality. A typical essay:</p> <ul style="list-style-type: none"> <li>• Effectively develops a point of view on the issue and demonstrates strong critical thinking, generally using appropriate examples, reasons and other evidence to support its position</li> <li>• Is well organized and focused, demonstrating coherence and progression of ideas</li> <li>• Exhibits facility in the use of language, using appropriate vocabulary</li> <li>• Demonstrates variety in sentence structure</li> <li>• Is generally free of most errors in grammar, usage and mechanics</li> </ul>
4	<p>An essay in this category demonstrates adequate mastery, although it has lapses in quality. A typical essay:</p> <ul style="list-style-type: none"> <li>• Develops a point of view on the issue and demonstrates competent critical thinking, using adequate examples, reasons and other evidence to support its position</li> <li>• Is generally organized and focused, demonstrating some coherence and progression of ideas</li> <li>• Exhibits adequate but inconsistent facility in the use of language, using generally appropriate vocabulary</li> <li>• Demonstrates some variety in sentence structure</li> <li>• Has some errors in grammar, usage and mechanics</li> </ul>
3	<p>An essay in this category demonstrates developing mastery, and is marked by ONE OR MORE of the following weaknesses:</p> <ul style="list-style-type: none"> <li>• Develops a point of view on the issue, demonstrating some critical thinking, but may do so inconsistently or use inadequate examples, reasons or other evidence to support its position</li> <li>• Is limited in its organization or focus, or may demonstrate some lapses in coherence or progression of ideas</li> <li>• Displays developing facility in the use of language, but sometimes uses weak vocabulary or inappropriate word choice</li> <li>• Lacks variety or demonstrates problems in sentence structure</li> <li>• Contains an accumulation of errors in grammar, usage and mechanics</li> </ul>
2	<p>An essay in this category demonstrates little mastery, and is flawed by ONE OR MORE of the following weaknesses:</p> <ul style="list-style-type: none"> <li>• Develops a point of view on the issue that is vague or seriously limited, and demonstrates weak critical thinking, providing inappropriate or insufficient examples, reasons or other evidence to support its position</li> <li>• Is poorly organized and/or focused, or demonstrates serious problems with coherence or progression of ideas</li> <li>• Displays very little facility in the use of language, using very limited vocabulary or incorrect word choice</li> <li>• Demonstrates frequent problems in sentence structure</li> <li>• Contains errors in grammar, usage and mechanics so serious that meaning is somewhat obscured</li> </ul>
1	<p>An essay in this category demonstrates very little or no mastery, and is severely flawed by ONE OR MORE of the following weaknesses:</p> <ul style="list-style-type: none"> <li>• Develops no viable point of view on the issue, or provides little or no evidence to support its position</li> <li>• Is disorganized or unfocused, resulting in a disjointed or incoherent essay</li> <li>• Displays fundamental errors in vocabulary</li> <li>• Demonstrates severe flaws in sentence structure</li> <li>• Contains pervasive errors in grammar, usage or mechanics that persistently interfere with meaning</li> </ul>

The following rubric was obtained from [collegereadiness.collegeboard.org/sat-essay-scoring-before-march-2016](http://collegereadiness.collegeboard.org/sat-essay-scoring-before-march-2016)

### *Prior Reading Ability*

Prior reading ability was assessed using the Gates-MacGinitie Reading Test (GMRT; MacGinitie and MacGinitie 1989). The GMRT is comprised of 48 multiple-choice questions about 11 unique passages. Participants were given 20-min to complete the GMRT after which they were automatically moved onto another task. Each item was scored correct/incorrect (1/0) to produce a numerical score out of 48. GMRT scores were computed by dividing the number of correct answers by the total number of questions to produce a proportion correct score.

### *System Interactions*

Participants' actions in iSTART and the Writing Pal were logged to assess time spent viewing instructional videos and engaging in practice activities. These values vary for each participant because some students skip through videos and others rewind and rewatch videos. The number of times videos and games were used was also assessed. Some participants watched videos multiple times as they closed the video window before completing the checkpoint and subsequent tasks were locked until checkpoint at the end of the video was completed in an attempt to prevent participants from skipping necessary tasks.

### *Source-Based Essay Questions*

Participants completed one randomly assigned 40-min content-specific source-based essay task<sup>14</sup> during this experiment. Content-specific source-based writing tasks do not rely on the writer's ability to locate source material; rather they provide a set of sources to be used during writing. The source-based essay prompts were counter-balanced within condition to ensure equal prompt representation within each condition. Two prompts were utilized to control for potential prompt effects due to reasons such as topic familiarity. Participants utilized a webpage to access the source materials and typed their essays in Microsoft Word.

The participants were informed that they would spend their second session composing a source-based essay and were provided the following general instructions. *Today you will be writing a source-based essay. You will have 40 min to read the sources below and respond to the following prompt. [Insert 1 Prompt from below] Make sure that your argument is central, use the sources provided in the file links below to illustrate and support your reasoning. Avoid merely summarizing the sources. Indicate clearly which sources you are drawing from, whether through direct quotation, paraphrase or summary. You may cite sources as Source A, Source B, etc. or by using the descriptions in parentheses.*

Two prompts were selected from past Advanced Placement Tests of English Language and Composition (the synthesis essay section; available from the College Board at [APcentral.collegeboard.com](http://APcentral.collegeboard.com), 2011 and 2011 Form B; College Board 2011a, 2011b, 2011c, 2011d). These prompts are designed to measure students' ability to read and evaluate multiple sources and their ability to select appropriate sources (for their

<sup>14</sup> The source-based (synthesis) essay on the AP English Language and Composition Test is completed during the two-hour free response section. The suggested time for this essay is 40 min.



stance) and integrate them into a coherent essay. The evidence and explanations used in the essays are evaluated along with features of the writing itself such as, grammar, syntax, cohesion, and organization to produce a holistic score. Prompts from the AP English Language and Composition Test were selected to control for required content knowledge and ability to locate source material. The source-based writing section on the English Language and Composition Test is ideal for our purposes as it is designed to evaluate the test taker's ability to read and evaluate multiple sources, and to synthesize this information into well-reasoned and written essays. The prompts selected were from the spring and summer 2011 exams and focus on related topics, green living and the locavore movement. Utilizing both tests from the same year guarantees a greater level of equivalence than selecting prompts from different years and are designed by the College Board to be equivalent in depth and difficulty.<sup>15</sup> The prompts supplied a different number of sources (6 vs 7), however, as these prompts are from the same testing year, designed to be equivalent, and require the same minimum use of source material, prompts were expected to represent similar difficulty levels. Sources included excerpts from news articles, books, graphs, and comics. With the exception of the comic and graphs, the text for each source was between ½ and a whole page in length. A summary of the sources provided for each prompt are provided in Tables 3 and 4.

### **Green Living**

Green living (practices that promote the conservation and wise use of natural resources) has become a topic of discussion in many parts of the world today. With changes in the availability and cost of natural resources, many people are discussing whether conservation should be required of all citizens.

Carefully read the following six sources, including the introductory information for each source. Then synthesize from at least three of the sources and incorporate it into a coherent, well-written essay that develops a position on the extent to which government should be responsible for fostering green practices.

### **Locavores**

Locavores are people who have decided to eat locally grown or produced products as much as possible. With an eye to nutrition as well as sustainability (resource use that preserves the environment), the locavore movement has become widespread over the past decade.

Imagine that a community is considering organizing a locavore movement. Carefully read the following seven sources, including the introductory information for each source. Then synthesize information from at least three of the sources and incorporate it into a coherent, well-developed essay that identifies the key issues associated with the locavore movement and examines their implications for the community.

<sup>15</sup> For more information see <https://professionals.collegeboard.org/testing/ap/about/different>

These essays were scored using the question-specific scoring guide released by College Board, which yield holistic scores ranging from 0 to 9. Though the scoring guides are prompt-specific, all scoring guides for the AP English Language and Composition Test free-response section prompt the rater to attend to content development, organization, coherence, and fluency and control of Standard Written English. Essays are scored holistically and categorize writing into 4 descriptive categories, unsuccessful (a score of 1 and 2), inadequate (3 and 4), adequate (6 and 7), and effective (8 and 9). A score of 5 represents an essay that is equally adequate and inadequate and a score of 0 represents a response that only repeats the prompt. The

**Table 3** Source material for green living prompt

Source	Source description	Summary
Winters, Sevastian. "The Pros and Cons of the United States 'Going Green': Is Environmental Consciousness Really All Good?" <i>Associated Content</i> . Associated Content, Inc., 3 Aug. 2009. Web. 18 Aug. 2009.	An excerpt from an online article about the United States going green.	This passage discusses the need for a mind-set of environmental stewardship and the benefits and drawbacks of going green.
Webber, Alan M. "U.S. Could Learn a Thing or Two from Singapore." Editorial. <i>USA Today</i> . USA Today, 14 Aug. 2006. Web. 17 Aug 2009.	An excerpt from an online editorial in a national newspaper	This passage discusses the policies in Singapore surrounding car ownership and investment in public transportation.
Friedman, Thomas L. <i>Hot, Flat, and Crowded: Why We Need a Green Revolution- and How It Can Renew America</i> . New York: Farrar, 2008. Print.	An excerpt from a book about the need for a green revolution.	This passage discusses America's weakened ability and willingness to tackle the problem of global warming and how America would benefit from taking on the challenge.
Samuelson, Robert J. "Selling the Green Economy." <i>Washington Post</i> . The Washington Post Company, 27 Apr. 2009. Web. 18 Aug. 2009.	An excerpt from an online article in a national newspaper.	This passage discusses the projected and actual costs of going green, and impacts of global warming.
Rheault, Magali. "In Top Polluting Nations, Efforts to Live 'Green' Vary." <i>Gallup</i> . Gallup, Inc. 22 Apr. 2008. Web. 18 Aug. 2009.	An excerpt from an article on the results of polls on environmental awareness conducted in 2007	A graph is presented that displays the percent of respondents in the five countries (U.S., China, Russia, Japan, and India) that produce 54 % of the world's total carbon dioxide emissions who have reported specific green practices such as recycling and water conservation.
United States. Department of Energy. Office of Energy Efficiency and Renewable Energy. <i>Energy Savers Booklet: Tips on Saving Energy &amp; Money at Home</i> . 6 Aug. 2009. Web. 18 Aug. 2009.	An excerpt from a website published by the United States Department of Energy	This source provides information on how to save money going green by making small changes

**Table 4** Source material for locavore prompt

Source	Source description	Summary
Maiser, Jennifer. "10 Reasons to Eat Local Food." <i>Eat Local Challenge</i> . Eat Local Challenge, 8 Apr. 2006. Web. 16 Dec. 2009.	An article from a group weblog written by individuals interested in the benefits of eating food grown and produced locally.	This passage discusses 10 arguments for eating locally grown and produced food.
Smith, Alisa, and J. B. MacKinnon. <i>Plenty: One Man, One Woman, and a Raucous Year Eating Locally</i> . New York: Harmony, 2007. Print.	An excerpt from a book written by the creators of the 100-mile diet an experience in eating only foods grown and produced within a 100-mile radius.	This passage discusses that the argument for local eating based on nutritional value as a red herring and reasons for eating local.
McWilliams, James E. "On My Mind: The Locavore Myth." <i>Forbes.com</i> . Forbes, 15 Jul. 2009. Web. 16 Dec. 2009.	An excerpt from an online opinion article in a business magazine.	This passage discusses the benefits of eating global over local in terms of energy usage and economic impact.
Loder, Natasha, Elizabeth Finkel, Craig Meisner, and Pamela Ronald. "The Problem of What to Eat." <i>Conservation Magazine</i> . The Society for Conservation Biology, July-Sept. 2008. Web. 16 Dec. 2009.	A chart from an online article in an environmental magazine.	The chart presented, <i>Total Greenhouse Gas Emissions by Supply Chain Tier Associated with Household Food Consumption in the United States</i> . It presents the green gas emissions from transportation, production, and wholesale/retail for eight categories of food.
Gogi, Pallavi. "The Rise of the 'Locavore': How the Strengthening Local Food Movement in Towns Across the U.S. is Reshaping Farms and Food Retailing." <i>Bloomberg Businessweek</i> . Bloomberg, 20 May 2008. Web. 17 Dec. 2009.	An excerpt from an online article in a business magazine.	This passage discusses the locavore movement is reshaping the business of growing and supplying food and the increase in the number of small farms in the U.S.
Roberts, Paul. <i>The End of Food</i> . New York: Houghton Mifflin Harcourt, 2008. Print.	An excerpt from a book about the food industry.	This source discusses problems with reliance on food shipped from halfway around the world, and the difficulties in implementing locavorism.
Hallatt, Alex. "Arctic Circle." Comic strip. King Features Syndicate, Inc. 1 Sept. 2008. Web. 12 July, 2009.	A cartoon from an environmentally themed comic strip.	This comic depicts a discussion in an igloo about locavorism with one character defining the local supermarket as eating local.

rubrics for the source-based writing prompts specifically focus on the development of a position, the synthesis of sources, the evidence and explanations provided, targeting the level, completeness, and appropriateness of explanations, the sophistication and clarity of the argument and argument development, the link between the source material and the argument, and fluency and control of Standard Written English, including the extent to which lapses in grammar, diction and syntax are distracting and detract from the meaning. Essays with numerous distracting errors in grammar and mechanics cannot be

scored higher than a 2 (unsuccessful) and those referencing fewer than 3 sources cannot score above a 4.

For the present study, the source-based essays were rated using a modified version of the rubric provided by the Advanced Placement Exam. Because participants were not explicitly instructed on and did not receive training on how to cite source material, credit was given for any direct reference to the source material. Not requiring explicit sourcing (e.g., source B) was the only change made to the scoring rubric. If a writer explicitly used a source (e.g., talked about the car tax in Singapore) but did not cite it, they were still given credit for utilizing that source.

## Results

### Time between Sessions

As sessions were scheduled at the convenience of participants differences in delay as a function of condition were assessed. The number of days between sessions ranged from 1 to 16, with 97 % of participants completing the second session within the 3 days following the initial session, and 99 % completing the second session within 5 days. The participant with a 16-day delay was maintained in the data set as this participant did not receive any training (control group). No difference was observed in time between sessions as a function of condition,  $F(3, 171) = 0.46$ ,  $p = .71$ . A one-way ANOVA was used to assess the impact of delay on source-based essay score,  $F(1, 170) = 1.93$ ,  $p = .17$ .

### Initial Assessment of Skills

#### *Descriptive and Correlation Analysis*

The means and standard deviations for the pre-test assessments of persuasive writing and reading skill (GMRT) by condition are presented in Table 5. As expected, students'

**Table 5** Means and standard deviations on pretest, testing, and outcome measures by condition

	Control	iSTART	Writing-Pal	Blended
Pretest Essay Score	3.69 (0.80)	3.56 (0.74)	3.80 (0.64)	3.98 (0.78)
Pretest Reading Score	0.61 (0.18)	0.63 (0.21)	0.62 (0.22)	0.70 (0.19)
Total System Time	--	86 min 19 s (14 min 58 s)	68 min 19 s (15min2s)	88min24s (11min30s)
Training Time	--	26 min 29 s (5 min 43 s)	32 min 30s (9 min 34 s)	46 min 34 s (11 min 38 s)
Practice Time	--	59 min 50 s (16 min 20s)	36 min 57 s (9 min 49 s)	41 min 50s (10 min 53 s)
Source-Based Essay Score	3.60 (1.36)	3.44 (1.40)	3.83 (1.28)	4.51 (1.74)

pre-test scores on the persuasive essay assessment and the GMRT reading assessment were moderately correlated ( $r = .42$ ). The following sections describe differences as a function of condition for each pre-test measure to establish equivalence of conditions.

### *Persuasive Writing*

Prompt effects were assessed to examine potential differences as a function of prompt ( $n_{\text{images and impression}} = 73$ ;  $n_{\text{competition and cooperation}} = 102$ ; see Table 6 for a breakdown of prompt assignment by condition), and no prompt effect on score was observed for participants,  $F(1, 173) = 3.01$ ,  $p = .085$ . Differences in initial writing ability were also assessed as a function of condition to assess equivalence between the groups. A non-significant trend for differences in initial writing ability was observed as a function of condition,  $F(3, 171) = 2.44$ ,  $p = .067$ , with those in the combined condition ( $M = 3.98$ ,  $SD = 0.78$ ) scoring slightly higher than participants in the iSTART condition ( $M = 3.56$ ,  $SD = 0.74$ ), with no differences in comparison to the control condition ( $M = 3.69$ ,  $SD = 0.80$ ) and the Writing Pal condition ( $M = 3.80$ ,  $SD = 0.64$ ). Prior writing ability will be included as a covariate in the main analysis to control for the impact of prior writing ability on content-specific source-based writing.

### *Gates MacGinitie Reading Test*

Differences in prior reading ability as a function of condition were also assessed. No differences in pretest reading ability were observed as a function of condition,  $F(3, 171) = 1.96$ ,  $p = .12$  ( $M_{\text{Control}} = 0.61$ ,  $SD = 0.18$ ;  $M_{\text{iSTART}} = 0.63$ ,  $SD = 0.21$ ;  $M_{\text{Writing-Pal}} = 0.62$ ,  $SD = 0.22$ ;  $M_{\text{Combined}} = 0.70$ ,  $SD = 0.19$ ). Prior reading ability will be included as a covariate in the main analysis to control for the impact of prior reading ability on content-specific source-based writing.

## **Performance during Training**

### *Training Time*

Aggregate times spent in a system, in training, and in practice are presented in Table 7. Total time spent in each system varied as a function of condition,  $F(2, 122) = 24.57$ ,  $p < .001$ . Those in the Writing Pal condition spent less total time interacting with the system than those in the iSTART and combined conditions. As such, the impact of strategy instruction as a function of training time is assessed in the following analyses and included as a covariate in the model.

**Table 6** Distribution of pretest prompts by condition

	Control	iSTART	Writing-Pal	Blended
Pretest Prompt				
Images and Impressions	21	15	18	19
Competition and Cooperation	27	26	23	26

**Table 7** Aggregate time spent within tutoring environments

Condition	Total System Time <i>M</i> ( <i>SD</i> )	Total Practice Time <i>M</i> ( <i>SD</i> )	Total Instruction Time <i>M</i> ( <i>SD</i> )
iSTART	86 min 19 s (14min58s)	59 min 50 s (16 min 20s)	26 min 29 s (5 min 43 s)
Writing Pal	68 min 19 s (15min2s)	36 min 57 s (9 min 49 s)	32 min 30s (9 min 34 s)
Blended	88 min 24 s (11min30s)	41 min 50s (10 min 53 s)	46 min 34 s (11 min 38 s)

**iSTART** An iSTART video was considered completed by a student if the time spent was within 10 s of the experimenter computed minimum time to finish the video. Some participants viewed the instructional videos multiple times during the study, primarily because participants closed videos early and had to view the video again (at least part of it) to trigger the checkpoint. Completion rates for the videos ranged from 66 % (Elaboration and Bridging) to 92 % (Overview and Demonstration). In total, participants received credit for watching between 1 and 8 videos with over half of the participants (58 %) watching all 8 videos. Unfortunately, over 20 % of participants completed fewer than half of the assigned videos. There was no difference in the number of videos completed as a function of condition (iSTART vs. combined),  $F(1, 81) = 2.37, p = .13$ . An analysis of overall instructional time in iSTART was possible as all participants interacting with iSTART were assigned to watch the same videos. Overall instruction time in iSTART ranged from 4 min 5 s – 50 min 53 s, with an average time of 12 min 16 s ( $SD = 13$  min 11 s). These numbers are skewed by those who did not watch videos; for those watching more than half of the videos, total instruction time ranged from 20 min 57 s – 50 min 53 s with an average instructional time of 27 min 20s ( $SD = 3$  min 32 s). There was a significant difference in instruction time as a function of condition,  $F(1, 81) = 3.98, p = .049$ , with those in the iSTART condition receiving on average almost 2 ½ minutes more instruction ( $M = 26$  min 29 s,  $SD = 5$  min 43 s) than those in the combined condition ( $M = 24$  min 2 s,  $SD = 5$  min 26 s). Average checkpoint performance ranged from 0 to 4 out of 4 possible points with an average score of 3.16 ( $SD = .78$ ). There was no difference in average checkpoint scores as a function of condition,  $F(1, 81) = .77, p = .38$ .

Conditions were designed so that all participants completed at least one text in Coached Practice; however three participants in the combined condition ran out of time and did not complete Coached Practice. As those in the iSTART condition were provided more time for coached practice and allowing for the potential completion of multiple texts only the average score from the first coached practice text is assessed here. For the first Coached Practice text, average self-explanation scores ranged from 1.61 to 3.00 with a mean of 2.57 ( $SD = .40$ ). Average self-explanation score did not vary as a function of condition,  $F(1, 80) = 2.17, p = .144$ . Participants spent between 6 min 8 s and 34 min 21 s completing the first text, with an average time of 14 min 17 s ( $SD = 5$  min 25 s). Participants spent on average 3 ½ minutes longer completing their first Coached Practice text if they were in the iSTART condition ( $M = 16$  min 2 s,  $SD = 6$  min 12 s), than if they were in the combined condition ( $M = 12$  min 37 s,

$SD = 3$  min 57 s),  $F(1, 82) = 8.93$ ,  $p = .004$ . Participants in the iSTART condition spent 45 min interacting with Coached Practice and 26 of those participants interacted with additional texts during that time. During the total 45 min participants interacted with from 1 to 5 texts with an average of 2.12 ( $SD = 1.12$ ) texts viewed. The average score across all Coached Practice texts for iSTART participants was 2.54 ( $SD = .38$ ).

After completing their first assigned time/text in Coached Practice participants had a variety of games available to them along with the continued availability of Coached Practice. Not all participants in the combined condition had time to complete games because the time to complete Writing Pal, the iSTART videos, and Coached Practice varied by participant. For those in the iSTART condition, only two continued to interact with Coached Practice instead of interacting with games. Two generative practice games were available to participants, 9 participants played Map Quest ( $M_{SE\ score} = 1.30$ ,  $SD = .48$ ) and 8 played Showdown ( $M_{SE\ score} = 2.05$ ,  $SD = .57$ ). There were no differences in average scores on any games observed as a function of condition.

**The Writing Pal** A Writing Pal video was considered completed by a student if the time spent was within 10 s of the minimum of the total video time. As in iSTART, some participants viewed the instructional videos multiple times during the study, primarily because participants closed videos early and had to view the video again (at least part of it) to trigger the checkpoint. Across conditions, completion rates for the videos ranged from 47 % (Summarize the Essay) to 89 % (Positions, Arguments, and Evidence). For the Writing Pal condition only, 40 % of participants watched all of the videos; similarly, only 39 % of participants in the combined condition watched all of the videos; over 20 % of Writing Pal participants and 38 % of combined participants completed fewer than half of the assigned videos. Because participants were assigned differing numbers of videos and game plays based on condition, direct comparisons of instructional time spent and game plays cannot be completed.

Performance scores on checkpoints in The Writing Pal were converted to proportion correct because checkpoints differed in number of questions. Average checkpoint proportion correct scores ranged from 22 to 95 % correct, with a mean of 74 % ( $SD = 14$  %). The full ranges of scores were observed for all checkpoints except those for Topic Sentences and Strengthening Your Evidence, for these checkpoints no participant received a score of zero. There was a marginal difference in checkpoint performance based on condition,  $F(1, 82) = 3.02$ ,  $p = .086$ , with those in the Writing Pal condition scoring on average of 5 % lower ( $M = 0.71$ ,  $SD = 0.15$ ) than those in the combined condition ( $M = 0.77$ ,  $SD = 0.13$ ). No differences were observed in game scores as a function of condition.

### Source-Based Essay Scores

Reliability in source-based essay scoring was established on 20 % of the essays. Adjacent accuracy (scores within one point) between the two raters<sup>16</sup> was 85.7 %, with

<sup>16</sup> The raters were experts who had conducted work in areas related to discourse studies and writing. There were not substantial differences in the knowledge and experience of the raters.

31.4 % exact agreement. This level of agreement is consistent with human scores for prompt-based persuasive essays where exact agreement ranges from 30 to 60 % and adjacent agreement from 85 to 100 % (Attali and Burstein 2006; Rudner et al. 2006; Shermis et al. 2010). Given comparative complexity of scoring source-based essays, this level of reliability was acceptable.

Participants wrote varying amounts, with the length of essays ranging from 116 words to 1036 words ( $M = 484.20$ ,  $SD = 167.89$ ). Across prompts, writers used an average of 2.86 sources out of either 6 or 7 sources (varied by prompt). Only one participant utilized all of the sources available, and three writers did not explicitly reference any source material. The source-based essays ranged in score from 1 to 8 (on a 0–9 scale) with a mean score of 3.85 ( $SD = 1.5$ ).

A total of 91 participants wrote essays on Locavorism and 84 wrote essays on Green Living (for prompt assignment by condition see Table 8). Prompts were randomly assigned to participant numbers prior to beginning the study to ensure that an equal number of responses were obtained for each prompt. However, as this study only examines the essays for participants whose first language is English, the assignments are not equal. Scores, word counts, and source frequency were assessed as a function of prompt to ensure that no differences existed due to the assigned topic and sources. No effect of prompt was found on score,  $F(1, 172) = .728$ ,  $p = .39$  or word count,  $F(1, 171) = 1.88$ ,  $p = .17$ . However, a significant difference in number of sources utilized was observed between the prompts,  $F(1, 172) = 5.59$ ,  $p = .02$ , with those responding to the prompt on green living utilizing more sources ( $M = 3.05$ ,  $SD = 1.04$ ) than participants responding to the locavore prompt ( $M = 2.69$ ,  $SD = .98$ ).

### Effects of Strategy Instruction

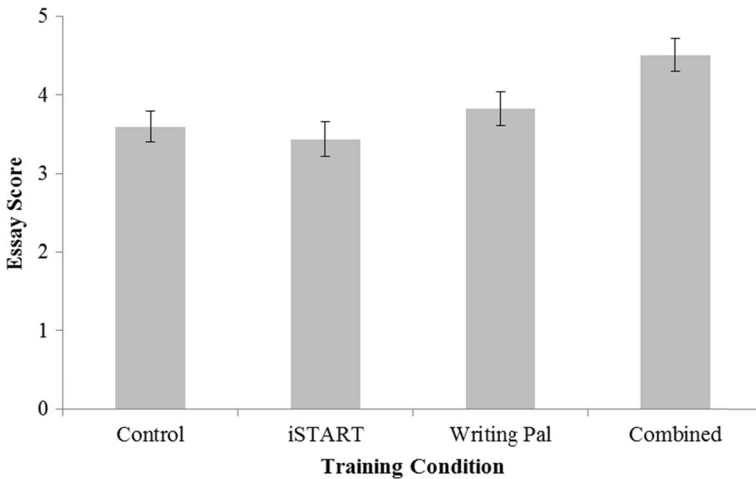
Half of the combined condition received each order of instruction ( $n_{iSTART-Writing Pal} = 23$ ,  $n_{Writing Pal - iSTART} = 22$ ). No difference was observed in source-based essay score as a function of the order of instruction,  $F(1, 43) = .002$ ,  $p = .97$ . Thus, all participants who received combined instruction were combined into a single group for all analyses. A one-way analysis of variance (ANOVA) was conducted to assess the impact of strategy instruction condition on source-based writing score. Performance on the source-based essay writing task varied as a function of type of strategy instruction completed,  $F(3, 171) = 4.61$ ,  $p = .004$ ,  $\eta^2 = .075$ . Post-hoc tests using Fisher's LSD test revealed that those in the combined instruction condition ( $M = 4.51$ ,  $SD = 1.75$ ) outperformed participants in the control ( $M = 3.60$ ,  $SD = 1.36$ ), iSTART ( $M = 3.44$ ,  $SD = 1.40$ ), and The Writing Pal ( $M = 3.83$ ,  $SD = 1.28$ ) conditions on source-based writing (see Fig. 1).

A second analysis using a one-way analysis of covariance (ANCOVA) was conducted to confirm that the impact of strategy instruction condition on source-based

**Table 8** Distribution of source-based essay prompt by condition

	Control	iSTART	Writing-Pal	Blended
Source-Based Essay Prompt				
Locavorism	25	21	22	23
Green Living	23	20	19	22





**Fig. 1** Source-based writing score by condition

essay score was not influenced by students' prior reading ability or writing proficiency. The covariates, writing proficiency [ $F(1, 169) = 5.51, p = .02, \eta^2 = .032; r = .31$ ] and reading ability [ $F(1, 169) = 7.71, p = .006, \eta^2 = .044; r = .32$ ] were significantly related to source-based essay score. However, as found in the previous analysis, the effect of strategy instruction condition remained significant,  $F(3, 169) = 2.17, p = .047, \eta^2 = .046$ . Additional analyses using hierarchical regression further confirmed that the impact of strategy instruction did not vary as a function of either reading or writing abilities (see Weston 2015, for details and additional analyses confirming the absence of training by aptitude interactions).

As training time varied as a function of condition the impact of training time was investigated. Nonetheless, the advantage for the combined condition remained when total training time was included within an ANCOVA. Focusing on the experimental conditions (i.e., there is no training for the control condition), a one-way ANCOVA was conducted to assess the impact of strategy instruction condition on source-based writing scores controlling for prior reading ability, writing proficiency, and total time spent on training. As found in the previous analyses, there were significant effects of strategy instruction condition,  $F(2, 120) = 3.17, p = .046, \eta^2 = .05$ , writing proficiency [ $F(1, 120) = 3.91, p = .05, \eta^2 = .032$ ] and reading ability [ $F(1, 120) = 4.88, p = .029, \eta^2 = .039$ ]; however, the covariate of total time spent on training was not significant [ $F(1, 120) = 0.57, p = .81, \eta^2 < .001; r = .065$ ]. The trends were equivalent when considering the impacts of practice time and instructional time separately.

In sum, combined training that included both reading and writing strategy training was effective regardless of prior reading and writing abilities.

## Discussion

Our overarching supposition of this study is that the production of high quality source-based essays is a complex task that relies on the development of *both* reading comprehension and writing skills. Our goal here was to address the gap in the literature

regarding the pedagogical techniques that most effectively improve students' ability to write content-specific source-based essays (scored both for quality of writing and content) by examining the effects of reading comprehension and writing strategy instruction and practice on content-specific source-based essay performance.

One method that has been proposed to impact both reading and writing skills is explicit strategy instruction (McNamara 2004; Roscoe and McNamara 2013). Such training has been shown to successfully improve performance on both reading and writing tasks (e.g., Graham and Harris 2007; McNamara 2007); yet, the combined impact of reading and writing strategy training has yet to be tested, particularly with respect to content-specific source-based essay writing.

In the current study, we capitalized on two ITSs, iSTART and the Writing Pal. Starting from the assumption that reading comprehension and writing are important elements of content-specific source-based essay writing, our aim was to assess the extent to which students' essay scores were impacted by receiving training that targeted these component skills. Strategy instruction and practice were provided to students in the context of computer-based learning environments. Specifically, the reading comprehension strategy instruction (iSTART) targeted self-explanation and reading comprehension strategies that are important for the comprehension of challenging texts. The writing strategy instruction (Writing Pal), on the other hand, focused on strategies for the three primary phases of the writing process (i.e., planning, drafting, revising). The students in this study were randomly assigned to receive training on reading, writing, or combined strategies, and the impact of these different forms of instruction was then examined.

We found that the combination of reading comprehension and writing strategy instruction positively impacted undergraduate students' performance on a timed, content-specific source-based essay in comparison to no training and either writing or comprehension strategy training alone. This is important because it indicates that the combination of reading and writing training can help to improve students' performance on source-based writing tasks. Further, the results revealed that, compared to a control condition, neither writing nor comprehension training significantly impacted performance on the source-based essay in absence of the other. This suggests that perhaps something "clicked" when the students were provided training on both processes, or that training in both domains primed the writer of the importance of both tasks, such that they were able to more successfully leverage the individual reading and writing strategies during the complex source-based writing task.

Follow-up analyses were conducted to examine the impact of individual differences and instructional time on the benefits observed from the combined strategy training. First, individual difference analyses revealed that the impact of the strategy training did not depend on students' literacy skills (i.e., reading and writing). This suggests, for example, that less skilled students benefitted from instruction as much as high skilled students. Importantly, these individual difference results may vary outside of the context of the current study. Here, we targeted undergraduate students, who were all (by chance) within a moderate range of abilities. Specifically, the majority of our participants scored in the average range on pre-test measures of reading and writing ability, with very few participants receiving scores indicative of very high or very low proficiency. This will not necessarily be the case in future studies, necessitating further investigation of the impact of prior reading and writing abilities. Indeed, previous

research with both iSTART and The Writing Pal has revealed differences in strategy training benefits based on individual differences, such as prior skill and knowledge, native language, and interest level (e.g., Allen et al. 2014a; Jackson et al. 2013). In iSTART and the Writing Pal, those with lower prior skill and knowledge generally benefit more than those with higher skill and knowledge (e.g., Jackson et al. 2010a, Jackson et al. 2013; McNamara 2004, 2016; McNamara et al. 2006). Additionally, research has revealed that individual differences among students influence their engagement in the systems, as well as the linguistic properties of their writing (Allen et al. 2014a; Allen et al. 2015). Hence, the effects of instruction should be expected to vary with different populations, particularly if their reading and writing instructional needs vary widely.

Results also indicated that the effects of instruction did not depend on time-on-task (i.e., overall time, instructional time, or practice time). Though time-on-task is often offered as an explanation for differences between training groups, the results of the current study suggest that overall instruction and practice time had no impact on source-based essay scores. This finding is important because it suggests that students in the combined condition were not negatively impacted by receiving less overall reading and writing strategy instruction, which further points to the multi-faceted nature of the source-based writing task.

## Conclusions

Educators and educational policies espouse the importance of writing; however, little has been done to improve instruction in this area. In particular, large class sizes and standardized testing demands have made it increasingly difficult for educators to adequately train students in writing (National Commission on Writing 2003). Further, many teachers report that they do not feel like they have the training necessary to teach writing (Leki 1990; Reid 1994; Susser 1994; Winer 1992). Given the lack of training and support teachers receive, it is important that researchers work to identify the most effective practices for improving writing proficiency.

The purpose of the current study was to examine differential effects of automated, adaptive strategy training on *content-specific source-based essay writing*. Source-based writing is a common means through which students' content knowledge is assessed in the classroom, and source-based essays are a commonly assigned in high school language arts classes and college classes (across disciplines). Content-specific source-based writing tasks, where the sources are provided to the writer, can be found on assessments across disciplines. However, little empirical research has been conducted on content-specific source-based essay writing compared to other tasks, such as persuasive writing. Additionally, the fact that source-based essay writing relies on literacy skills (i.e., reading and writing) beyond content knowledge is often overlooked. The results of our study emphasize the critical point that source-based essay writing is a developed *skill* that can be improved through systematic training and practice. In particular, the results suggest that source-based writing performance may be improved through a combination of reading and writing strategy training. It is important to note that these results only apply to content-specific source-based essay writing, not all source-based essay writing, as the task of finding and selecting appropriate source

material adds an additional level of complexity to the task. Additionally, these results may not transfer to source-based writing tasks where only content or writing proficiency were targeted and assessed.

One significant contribution of this study is its demonstration that students can gain from instruction in both reading and writing. Notably, however, we do not know *how* students gained from combined instruction in both reading and writing. These tutoring systems are both guided by principles of learning such as active retrieval, deliberate practice and feedback, and the inclusion of motivational elements (McNamara et al. 2015b), and they have undergone many years of testing. But we do not know whether any handful of literacy interventions might lead to similar gains on source-based writing tasks. We find the latter quite doubtful given the general consensus that helping students improve on this type of task is quite challenging, and given the focus of the Common Core and multiple national and international agencies on this common problem. Nonetheless, we do not yet know which are the key ingredients that led to students' gains. As such, this study points to many avenues of future research to further investigate ways to improve students' ability to compose source-based essays, and which elements of instruction comprise the key ingredients. Importantly, neither tutoring system used in this study provided instruction or practice directly targeting strategies for source-based writing. Hence, the results demonstrate that the combination of iSTART and Writing Pal training transferred to a far transfer task. Moreover, the two ITSs were provided to students in an ad-hoc fashion, without explanation as to why the systems might enhance their performance on source-based writing. This aspect of training emanated from the constraints imposed by the experimental study to examine the independent effects of reading and writing strategy training.

These observed benefits from the loosely aligned training suggest the possibility of even greater benefits if training were more closely aligned with the task. Such an alignment may only require minimal changes to the training. For example, in iSTART, two additional modules might be envisioned – one on the selection of relevant information for answering questions and a second module on the use of the bridging inference strategy to make connections between sources. Similarly, modules could be added to the Writing Pal related to the inclusion and selection of source material and on the strategies needed to effectively compare and contrast sources in writing. Perhaps most optimally, a combined system might be developed to provide training that systematically targets students' performance on source-based essays. We would expect the success of such a system to be further enhanced by providing students with opportunities for deliberate practice, which would include automated feedback that specifically focuses on source-based writing. This, in turn, depends on the development of computational algorithms to automatically score and provide feedback on these forms of essays.

Nonetheless, with minimal adjustments to iSTART and the Writing Pal, the combination of the two automated tutoring systems led to improvements in students' writing performance. Our overarching aim will be to move beyond this research design to more systematically investigate the skills underlying source-based writing tasks. This future research will allow us to better adapt the reading and writing strategy training to these writing tasks, which will ideally help students to improve their success in a wide variety of content domains.

**Acknowledgments** This research was supported by the Institute for Education Sciences (IES R305A120707; IES R305A130124). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES. The Authors would like to thank Cecile Perret for helping to collect the data, and Matt Jacovina and Amy Johnson for their help with scoring the essays.

## References

- Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014a). Game-based writing strategy tutoring for second language learners: game enjoyment as a key to engagement. *Language, Learning and Technology*, 18, 124–150.
- Allen, L. K., Snow, E. L., Crossley, S. A., Jackson, G. T., & McNamara, D. S. (2014b). Reading comprehension components and their relation to the writing process. *L'année psychologique/Topics in Cognitive Psychology*, 114, 663–691.
- Allen, L. K., Crossley, S. A., Snow, E. L., Jacovina, M. E., Perret, C. A., & McNamara, D. S. (2015). Am I wrong or am I right? Gains in monitoring accuracy in an intelligent tutoring system for writing. In A. Mitrovic, F. Verdejo, C. Conati, & N. Heffernan (Eds.), *Proceedings of the 17th international conference on artificial intelligence in education (AIED 2015)*, (pp. 533–536). Madrid.
- Allen, L. K., Perret, C. A., & McNamara, D. S. (2016). Linguistic signatures of cognitive processes during writing. In J. Trueswell, A. Papafragou, D. Grodner, & D. Mirman (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society in Philadelphia, PA.*, (pp. 2483–2488). Austin, TX: Cognitive Science Society.
- Anmarkrud, O., Bråten, I., & Strømsø, H. I. (2014). Multiple-documents literacy: strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences*, 30, 64–75.
- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater V.2. *Journal of Technology, Learning, and Assessment*, 4(3). Retrieved from [www.jtla.org](http://www.jtla.org).
- Beminger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading connections between language by hand and language by eye. *Journal of Learning Disabilities*, 35(1), 39–56.
- Braine, G. (1995). Writing in the natural sciences and engineering. In D. Belcher & G. S. Braine (Eds.), *Academic writing in a second language: essays on research and pedagogy* (pp. 113–134). Norwood: Albx.
- Britt, M. A., & Aglinskias, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, 20(4), 485–522.
- Brown, A. (1982). Learning how to learn from reading. In J. A. Langer & M. T. Smith-Burke (Eds.), *Reader meets author: bridging the gap* (pp. 26–54). Newark: International Reading Association.
- Cameron, A., & Moshenko, B. (1996). Elicitation of knowledge transformational reports while children write narratives. *Canadian Journal of Behavioural Science*, 2, 271–280.
- College Board. (2011a). AP English Language and Composition 2011 Free-Response Questions. Retrieved from [http://apcentral.collegeboard.com/apc/public/repository/ap11\\_frq\\_english\\_language.pdf](http://apcentral.collegeboard.com/apc/public/repository/ap11_frq_english_language.pdf).
- College Board (2011b). AP English Language and Composition 2011 Scoring Guidelines. Retrieved from [http://apcentral.collegeboard.com/apc/public/repository/ap11\\_frq\\_english\\_language.pdf](http://apcentral.collegeboard.com/apc/public/repository/ap11_frq_english_language.pdf).
- College Board (2011c) AP English Language and Composition 2011 Free-Response Questions Form B. Retrieved from [http://apcentral.collegeboard.com/apc/public/repository/ap11\\_frq\\_english\\_language\\_formb.pdf](http://apcentral.collegeboard.com/apc/public/repository/ap11_frq_english_language_formb.pdf).
- College Board (2011d) AP English Language and Composition 2011 Scoring Guidelines Form B. Retrieved from [http://apcentral.collegeboard.com/apc/public/repository/ap11\\_english\\_language\\_form\\_b\\_scoring\\_guidelines.pdf](http://apcentral.collegeboard.com/apc/public/repository/ap11_english_language_form_b_scoring_guidelines.pdf).
- Crossley, S. A., Weston, J., McLain-Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: a linguistic analysis. *Written Communication*, 28, 282–311.
- Crossley, S. A., Roscoe, R. D., & McNamara, D. S. (2013). Using automatic scoring models to detect changes in student writing in an intelligent tutoring system. In C. Boonthum-Denecke & G. M. Youngblood (Eds.), *Proceedings of the 26th annual Florida artificial intelligence research society (FLAIRS) conference* (pp. 208–213). Menlo Park: The AAAI Press.
- Diederich, P. (1966). How to measure growth in writing ability. *English Journal*, 55, 435–449.

- Eblan, C. (1983). Writing across the curriculum: a survey of a university faculty's views and classroom practices. *Research in the Teaching of English*, 17, 343–348.
- Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, 32, 400–414.
- Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, 35, 39–50. doi:10.1207/S15326985EP3501\_5.
- Flower, L. S., & Hayes, J. R. (1980). The cognition of discovery: defining a rhetorical problem. *College Composition and Communication*, 31, 21–32.
- Foltz, P. W., Britt, M. A., & Perfetti, C. A. (1996). Reasoning from multiple texts: an automatic analysis of readers' situation models. In *Proceedings of the 18th annual cognitive science conference* (pp. 110–115). New Jersey: Lawrence Erlbaum.
- Gerard, L. F., Ryoo, K., McElhaney, K. W., Liu, O. L., Rafferty, A. N., & Linn, M. C. (2016). Automated guidance for student inquiry. *Journal of Educational Psychology*, 108, 60–81.
- Graesser, A. C., McNamara, D. S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36, 193–202.
- Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist*, 35, 3–12.
- Graham, S., & Harris, K. R. (2007). Best practices in teaching planning. Best practices in writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (pp. 119–140). New York: Guilford.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, 445–476.
- Hebert, M., Gillespie, A., & Graham, S. (2013). Comparing effects of different writing activities on reading comprehension: a meta-analysis. *Reading and Writing*, 26(1), 116–138. doi:10.1007/s11145-012-9386-3.
- Henry, A., & Roseberry, R. L. (1997). An investigation of the functions, strategies and linguistic features of the introductions and conclusions of essays. *System*, 24, 479–495.
- Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology*, 105, 1036–1049.
- Jackson, G. T., Boonthum, C., & McNamara, D. S. (2010a). The efficacy of iSTART extended practice: low ability students catch up. In J. Kay & V. Aleven (Eds.), *Proceedings of the 10th international conference on intelligent tutoring systems* (pp. 349–351). Berlin/Heidelberg: Springer.
- Jackson, G. T., Guess, R. H., & McNamara, D. S. (2010b). Assessing cognitively complex strategy use in an untrained domain. *Topics in Cognitive Science*, 2, 127–137.
- Jackson, G. T., Varner (Allen), L. K., Boonthum-Denecke, C., & McNamara, D. S. (2013). The impact of individual differences on learning with an educational game and a traditional ITS. *International Journal of Learning Technology*, 8, 315–336.
- Juel, C., Griffith, P., & Gough, P. (1986). A longitudinal study of the changing relationships of word recognition, spelling, reading comprehension, and writing from first to second grade. *Journal of Educational Psychology*, 78, 243–255.
- Landauer, T. K., McNamara, D. S., Dennis, S. E., & Kintsch, W. E. (2007). *Handbook of latent semantic analysis*. Mahwah: Lawrence Erlbaum Associates Publishers.
- Leki, I. (1990). Coaching from the margins: issues in written response. In B. Kroll (Ed.), *Second language writing: research insights for the classroom* (pp. 155–177). Cambridge: Cambridge University Press.
- Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. *Science Education*, 87, 517–538.
- Loban, W. D. (1967). *Language ability—grades 10, 11, and 12: final report*. Urbana: National Council of Teachers of English.
- MacGinitie, W. H., & MacGinitie, R. K. (1989). *Gates MacGinitie reading tests*. Chicago: Riverside.
- Magliano, J. P., Todaro, S., Millis, K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S. (2005). Changes in reading strategies as a function of reading training: a comparison of live and computerized training. *Journal of Educational Computing Research*, 32, 185–208.
- McCarthy, P. M., Renner, A. M., Duncan, M. G., Duran, N. D., Lightman, E. J., & McNamara, D. S. (2008). Identifying topic sentencehood. *Behavior Research Methods*, 40, 647–664.
- McNamara, D. S. (2004). SERT: self-explanation reading training. *Discourse Processes*, 38(1), 1–30.
- McNamara, D. S. (Ed.). (2007). *Reading comprehension strategies: Theory, interventions, and technologies*. Mahwah, NJ: Erlbaum.
- McNamara, D. S. (2013). The epistemic stance between the author and reader: a driving force in the cohesion of text and writing. *Discourse Studies*, 15(5), 1–17.

- McNamara, D. S. (2015). Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. *Discourse Processes*. doi:10.1080/0163853X.2015.1101328. Advance online publication
- McNamara, D. S. (2016). Self-Explanation and Reading Strategy Training (SERT) Improves low-knowledge students' science course performance. *Discourse Processes* (in press).
- McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: an automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing and content analysis: identification, investigation, and resolution* (pp. 188–205). Hershey: IGI Global.
- McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: interactive strategy trainer for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, 36, 222–233.
- McNamara, D. S., O'Reilly, T. P., Best, R. M., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, 34, 147–171.
- McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I. B. (2007a). iSTART: a web-based tutor that teaches self-explanation and metacognitive reading strategies. In D. S. McNamara (Ed.), *Reading comprehension strategies: theories, interventions, and technologies* (pp. 397–420). Mahwah: Erlbaum.
- McNamara, D. S., Ozuru, Y., Best, R., & O'Reilly, T. (2007b). The 4-pronged comprehension strategy framework. In D. S. McNamara (Ed.), *Reading comprehension strategies: theories, interventions, and technologies* (pp. 465–496). Mahwah: Erlbaum.
- McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007c). Evaluating self-explanations in iSTART: comparing word-based and LSA algorithms. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 227–241). Mahwah: Lawrence Erlbaum Associates, Inc..
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27, 57–86.
- McNamara, D. S., Crossley, S. A., & Roscoe, R. D. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45, 499–515.
- McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.
- McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015a). Hierarchical classification approach to automated essay scoring. *Assessing Writing*, 23, 35–59.
- McNamara, D. S., Jacovina, M. E., Snow, E. L., & Allen, L. K. (2015b). From generating in the lab to tutoring systems in classrooms. *American Journal of Psychology*, 128(2), 159–172.
- National Assessment of Educational Progress. (2011). The Nation's Report Card: Writing 2011. Retrieved Nov. 5, 2012, nces.ed.gov/nationsreportcard/writing.
- National Assessment of Educational Progress. (2013). The Nation's Report Card: Reading 2013 Retrieved, nces.ed.gov/nationsreportcard/reading.
- National Commission on Writing (2003). *The Neglected "R."*. New York: College Entrance Examination Board.
- Newell, G. E. (2007). Writing to learn: how alternative theories of school writing account for student performance. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 235–247). New York: The Guilford Press.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117–175. doi:10.1207/s1532690xci0102\_1.
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic inquiry and word count: LIWC 2007*. Austin: LIWC.net www.liwc.net.
- Reid, J. (1994). Responding to ESL students' texts: the myths of appropriation. *TESOL Quarterly*, 28, 273–294.
- Roscoe, R. D., & McNamara, D. S. (2013). Writing pal: feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, 105, 1010–1025. doi:10.1037/a0032340.
- Roscoe, R. D., Snow, E. L., Brandon, R. D., & McNamara, D. S. (2013). Educational game enjoyment, perceptions, and features in an intelligent writing tutor. In C. Boonthum-Denecke & G. M. Youngblood (Eds.), *Proceedings of the 26th annual Florida artificial intelligence research society (FLAIRS) conference* (pp. 515–520). Menlo Park: The AAAI Press.
- Roscoe, R. D., Brandon, R. D., Snow, E. L., & McNamara, D. S. (2014). Game-based writing strategy practice with the writing Pal. In K. Pytash & R. Ferdig (Eds.), *Exploring technology for writing and writing instruction* (pp. 1–20). Hershey: IGI Global.
- Rouet, J. F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology*, 88(3), 478.

- Rudner, L., Garcia, V., & Welch, C. (2006). An evaluation of the IntelliMetric essay scoring system. *Journal of Technology Learning, and Assessment*, 4, 3–21.
- Shermis, M. D., Burstein, J., Higgins, D., & Zechner, K. (2010). Automated essay scoring: writing assessment and instruction. In E. Baker, B. McGaw, & N. S. Petersen (Eds.), *International encyclopedia of education*. Oxford: Elsevier.
- Snow, E. L., Jacovina, M. E., Jackson, G. T., & McNamara, D. S. (2016). iSTART-2: a reading comprehension and strategy instruction tutor. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive educational Technologies for Literacy Instruction*. New York: Taylor & Francis, Routledge.
- Spivey, N. N. (1990). Transforming texts constructive processes in reading and writing. *Written Communication*, 7, 256–287.
- Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: the role of meta-cognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning*, 2, 191–210.
- Susser, B. (1994). Process approaches in ESL/BFL writing instruction. *Journal of Second Language Writing*, 3(3), 1–47.
- Tierney, R. J., & Shanahan, T. (1991). Research on the reading-writing relationship: interactions, transactions, and outcomes. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *The handbook of reading research* (Vol. 2, pp. 246–280). New York: Longman.
- Weston, J. L. (2015). *The impact of strategy instruction on source-based essay writing* (Doctoral dissertation). Retrieved from Proquest. (3718880).
- Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal*, 46, 1060–1106.
- Wiley, J., Steffens, B., Britt, M. A. & Griffin, T. D. (2014). Writing to learn from multiple-source inquiry activities in history. In G. Rijlaarsdam (Series Ed.) and P. Klein, P. Boscolo, C. Gelati, & L. Kilpatrick (Volume Eds.), *Studies in writing, writing as a learning activity* (pp. 120–148). England: Emerald Group Publishing Limited.
- Wineburg, S. S. (1991). Historical problem solving: a study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73–87.
- Winer, L. (1992). Spinach to chocolate: changing awareness and attitudes in ESL writing teachers. *TESOL Quarterly*, 26, 57–79.
- Zimmerman, B., & Risemberg, R. (1997). Becoming a self-regulated writer: a social cognitive perspective. *Contemporary Educational Psychology*, 22, 73–101.